

1 **Investigating Sources of Variability and Error in Simulations of Carbon Dioxide in**
2 **an Urban Region**
3

4 **Cory R. Martin^a, N. Zeng^{a,b}, A. Karion^c, K. Mueller^c, S. Ghosh^d, I. Lopez-Coto^d, K. R.**
5 **Gurney^e, T. Oda^{f,g}, K. Prasad^d, Y. Liu^{a*}, R. R. Dickerson^{a,b}, and J. Whetstone^c**

6 ^aDepartment of Atmospheric and Oceanic Science, University of Maryland, College Park, MD,
7 USA

8 ^bEarth System Science Interdisciplinary Center, University of Maryland, College Park, MD,
9 USA

10 ^cSpecial Programs Office, National Institute of Standards and Technology, Gaithersburg, MD,
11 USA

12 ^dFire Research Division, National Institute of Standards and Technology, Gaithersburg, MD,
13 USA

14 ^eSchool of Life Sciences, Arizona State University, Tempe, AZ, USA

15 ^fGlobal Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD,
16 USA

17 ^gGoddard Earth Sciences Research and Technology, Universities Space Research Association,
18 Columbia, MD, USA

19 ^{*}Now at: College of Geosciences, Texas A&M University, College Station, TX, USA

20 Corresponding authors: Cory Martin (cmart90@umd.edu); Ning Zeng (zeng@umd.edu)

21 Ning Zeng

22 2417 Atlantic Building

23 University of Maryland, College Park

24 College Park, MD 20742

25 301-405-5377

26 Fax: 301-314-9482

27
28 Declarations of interest: none

29 **Highlights:**

- 30 • Evaluation of modeled urban carbon dioxide using multiple emissions inventories
31 • Modeled carbon dioxide mole fractions agree with observations on average within 1%
32 • Spread in emissions inventories secondary to error resulting from model meteorology
33 • Synoptic meteorology as important as time of day for simulating observations
34

35 **Abstract**

36 As cities embark upon greenhouse gas (GHG) mitigation efforts, there is an increasing need for
37 accurate quantification of urban emissions. In urban areas, transport and dispersion is particularly
38 difficult to simulate using current mesoscale meteorological models due, in part, to added
39 complexity from surface heterogeneity and fine spatial/temporal scales. It is generally assumed
40 that the errors in GHG estimation methods in urban areas are dominated by errors in transport
41 and dispersion. Other significant errors include, but are not limited to, those from assumed
42 emissions magnitude and spatial distribution. To assess the predictability of simulated trace gas
43 mole fractions in urban observing systems using a numerical weather prediction model, we
44 employ an Eulerian model that combines traditional meteorological variables with multiple
45 passive tracers of atmospheric carbon dioxide (CO₂) from anthropogenic inventories and a
46 biospheric model. The predictability of the Eulerian model is assessed by comparing simulated
47 atmospheric CO₂ mole fractions to observations from four in situ tower sites (three urban and
48 one rural) in the Washington DC/Baltimore, MD area for February 2016. Four different gridded
49 fossil fuel emissions inventories along with a biospheric flux model are used to create an
50 ensemble of simulated atmospheric CO₂ observations within the model. These ensembles help to
51 evaluate whether the modeled observations are impacted more by the underlying emissions or
52 transport. The spread of modeled observations using the four emission fields indicates the
53 model's ability to distinguish between the different inventories under various meteorological
54 conditions. Overall, the Eulerian model performs well; simulated and observed average CO₂
55 mole fractions agree within 1% when averaged at the three urban sites across the month.
56 However, there can be differences greater than 10% at any given hour, which are attributed to
57 complex meteorological conditions rather than differences in the inventories themselves. On
58 average, the mean absolute error of the simulated compared to actual observations is generally
59 twice as large as the standard deviation of the modeled mole fractions across the four emission
60 inventories. This result supports the assumption, in urban domains, that the predicted mole
61 fraction error relative to observations is dominated by errors in model meteorology rather than
62 errors in the underlying fluxes in winter months. As such, minimizing errors associated with
63 atmospheric transport and dispersion may help improve the performance of GHG estimation
64 models more so than improving flux priors in the winter months. We also find that the errors
65 associated with atmospheric transport in urban domains are not restricted to certain times of day.
66 This suggests that atmospheric inversions should use CO₂ observations that have been filtered
67 using meteorological observations rather than assuming that meteorological modeling is most
68 accurate at certain times of day (such as using only mid-afternoon observations).

69

70 **Key Words**

71 Urban, climate, flux, emissions, CO₂

72

73 **1 Introduction**

74 Anthropogenic activities such as the combustion of fossil fuels contribute to the largest net flux
75 of carbon into the atmosphere (Intergovernmental Panel on Climate Change, 2013). These
76 emissions along with fluxes associated with deforestation have caused globally averaged CO₂
77 concentrations to rise from approximately 280 $\mu\text{mol mol}^{-1}$ of dry air (parts per million, or ppm)

78 at the beginning of the preindustrial era (Etheridge et al., 1996), to a level today of over 400 ppm
79 (NOAA, 2018), a value that continues to rise year after year. Urban areas contribute roughly
80 70% of the global total fossil fuel emissions (United Nations, 2011). This large contribution of
81 CO₂ emissions from urban areas is due to the generation of electrical and industrial energy
82 generation as well as vehicular transportation, among other things (United Nations, 2011). For
83 regional or global scales, the uncertainty of fossil fuel CO₂ emissions is assumed to be small
84 relative to that of the natural fluxes, but this assumption likely does not hold for smaller scales as
85 biases in both spatial patterns and magnitudes may occur (e.g., Hutchins et al. 2016; Fischer et
86 al., 2017; Gately and Hutyra 2017; Oda et al. 2018) As cities embark upon GHG mitigation
87 efforts, accurate quantification of urban emissions may be able to significantly inform their
88 effective management, in addition to providing quantitative substantiation of progress toward
89 emission reduction goals, many of which extend well into the 21st century (Clarke et al., 2014).

90 As such, in recent years, several urban GHG measurement campaigns have been implemented to
91 improve measurement, quantification, and attribution of fossil fuel CO₂ emission fluxes in
92 metropolitan areas along with their associated uncertainties. These cities include, but are not
93 limited to, Salt Lake City (McKain et al., 2012), Boston (Briber et al., 2013; Sargent et al.,
94 2018), Indianapolis (Turnbull et al., 2015; Lauvaux et al., 2016; Miles et al., 2017), Paris (Breón
95 et al., 2015), Davos (Lauvaux et al., 2013), and Los Angeles (Kort et al., 2013; Feng et al.,
96 2016). Generally, these experiments feature a network of towers equipped with state-of-the-art
97 greenhouse gas analyzers with mole fraction standards, however other networks with differing
98 observation systems exist (Wunch et al., 2010; Strong et al., 2011; Shusterman et al., 2016;
99 Vogel et al., under review).

100 The observations from towers upwind, downwind, and inside the metropolitan areas are used in
101 conjunction with atmospheric transport and Lagrangian dispersion models and with prior
102 emission distributions to estimate CO₂ fluxes (both spatially and temporally resolved) and
103 associated uncertainties using a Bayesian approach. These statistical methods are known as
104 inverse models and are commonly considered a “top down” approach since the estimated flux is
105 calculated using atmospheric observations. Urban emissions can also be derived using emission
106 factors applied to emission proxies such as buildings, traffic counts, and power plants (aka
107 “bottom-up” method; Ciais et al., 2010; Leip et al., 2018) that yield anthropogenic flux
108 inventories (e.g., Gurney et al., 2009; Oda and Maksyutov, 2011; McDonald et al., 2014; Gateley
109 and Hutyra, 2017). Theoretically, these two approaches could be completely independent from
110 one another. However, given limited atmospheric constraint along with the diffuse nature of the
111 atmosphere, inventories are generally used as a priori information of GHG emissions to help
112 constrain spatial patterns of emission estimates. Thus, transport and dispersion models and the
113 representation of the underlying flux distribution play a large role in accurately estimating GHG
114 emissions at the necessary spatial and temporal scales for mitigation purposes. Although both
115 types of errors (in modeled fluxes and transport) in urban domains have been explored (e.g.,
116 Deng et al., 2017), relative magnitudes and relationships continue to be uncertain especially at
117 observational time scales.

118 One of the most recently established urban GHG observation networks is the National Institute
119 of Standards and Technology (NIST) Northeast Corridor (NEC-B/W), currently encompassing
120 the Baltimore, Maryland and Washington, DC metropolitan areas (Lopez-Coto et al., 2017;
121 Mueller et al., 2018). This network was implemented to demonstrate and improve measurement
122 capabilities for quantifying anthropogenic GHG emissions from urban areas that cannot easily be

123 disentangled from one another. It is expected that meteorological conditions are not spatially
124 uniform across the region and are temporally impacted by distinct synoptic events. Although
125 this campaign will ultimately consist of a sixteen-tower network, in 2016 only three towers were
126 operational. The observations from these three towers provide an opportunity to (1) investigate
127 the ability to predict mole fractions using an atmospheric transport and dispersion model, along
128 with prior flux distributions and (2) assess the relative impact of transport and prior errors on the
129 simulated observations.

130 To achieve these objectives, we employ an Eulerian transport model that includes passive
131 chemical tracers that use emission inventories as the surface flux along with initial and boundary
132 conditions to generate 4D fields of atmospheric CO₂. In this way, we can vary the tracers to
133 examine the sensitivity of the predicted mole fractions compared to the assumed meteorology.
134 Eulerian models advect and disperse GHGs forward in time (e.g. Feng et al., 2016) compared to
135 Lagrangian approaches that use particle dispersion models operating backward from an
136 observational 4D location (e.g. McKain et al., 2012). These are analogous approaches, but we
137 employ a Eulerian model so that we can examine simulated meteorology for the entire domain to
138 help us interpret model performance (at added computational cost compared to a Lagrangian
139 model). In addition, we focus on CO₂ given the availability of data, specifically inventory data so
140 that multiple inventories can be used to estimate errors resulting from emissions inventories.
141 Simulated CO₂ mole fractions are compared to CO₂ observations from four in situ towers sites
142 (three urban and one rural) in the NEC-B/W for the month of February 2016.

143 The paper is outlined as follows: Section 2 describes the methods used for this analysis,
144 including the model domain and configuration, the observation datasets used in the evaluation,
145 and the emissions inventories as input to the transport model, Section 3 presents the observed
146 and modeled CO₂ time series at specific locations, and compares the model meteorology and
147 predicted mole fractions to observed values. A discussion of the results is featured in Section 4
148 and a summary of results and conclusions are presented in Section 5.

149 **2 Methods**

150 The main component of the modeling framework described in this study is the Weather Research
151 and Forecasting model coupled with chemistry (WRF-Chem), a non-hydrostatic, compressible
152 model that provides passive tracer transport online with mesoscale meteorology forecasting
153 capabilities (Grell et al., 2005; Skamarock et al., 2008; Beck et al., 2011). WRF-Chem has been
154 modified to allow for separate passive CO₂ tracers for four anthropogenic emissions inventories.
155 To evaluate the modeled CO₂'s sensitivity to the tracer input, we employ inventories that are
156 commonly used as prior anthropogenic fluxes in inverse modeling studies. In this paper, we refer
157 to a tracer as a 4D mole fraction field of CO₂ whereas the emissions inventory refers to the 3D
158 (or 2D if it does not have temporal variability) flux field. Additionally, a tracer for the biogenic
159 component of the CO₂ concentrations is also included in this modified version of WRF-Chem
160 since the mole fractions observed at tower locations are the integrated signal of both biospheric
161 and anthropogenic fluxes on top of the global atmospheric concentration. For the subsequent
162 analysis presented in this paper, WRF-Chem was run for the month of February 2016. The
163 month of February is used because it is assumed that anthropogenic emissions dominate the
164 integrated atmospheric signals as observed from these tower locations during winter months, as
165 wintertime fluxes from the biosphere are assumed to be smaller relative to anthropogenic
166 emissions, which is not the case during summer months. Thus, the impact of differences between

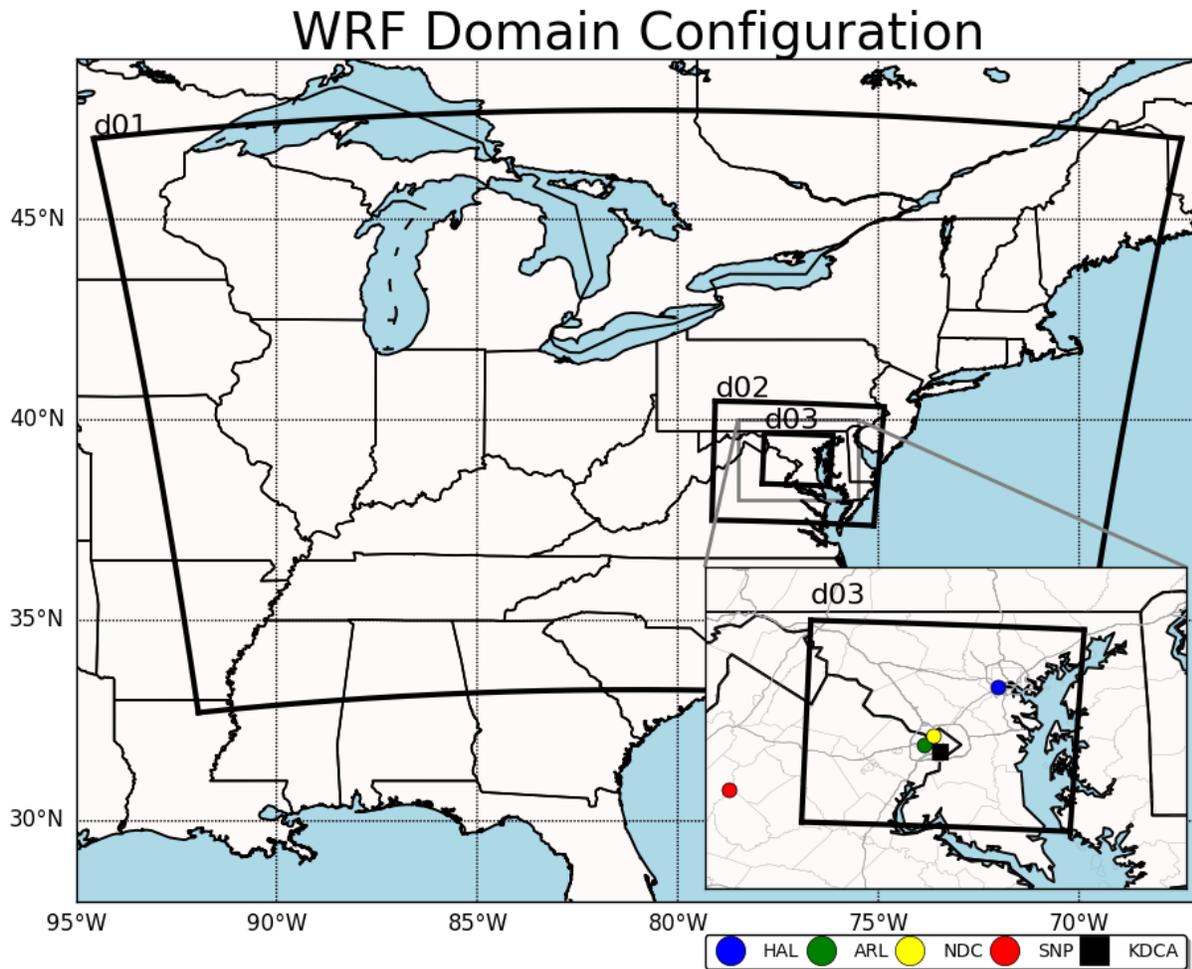
167 various prior anthropogenic flux estimates can be better ascertained in the simulated
168 observations. Future work will include evaluating the biospheric model by simulating a
169 summertime period.

170 **2.1 CO₂ Observations**

171 The NEC-B/W will ultimately feature a network of 16 observation sites (12 urban/suburban sites
172 and 4 rural sites) to measure CO₂ and methane (CH₄) continuously. Operated by Earth Networks
173 (EN), each site will feature a high-precision CRDS greenhouse gas analyzer and a calibration
174 and data processing system similar to the in situ sites in the Los Angeles Megacities project
175 (Verhulst et al., 2017). Additionally, a software-controlled valve system switches between
176 multiple inlets, where available, to pull ambient air to sample from different heights above
177 ground level. Data are quality controlled and averaged to hourly mole fractions reported on the
178 WMO X2007 scale (CO₂; Zhao and Tans, 2006) and X2004A scale (CH₄; Dlugokencky et al.,
179 2005) for each inlet height.

180 In February 2016, three GHG observation tower sites had been established and were collecting
181 continuous in situ CO₂ mole fraction measurements. The three sites are: HAL in Halethorpe, MD
182 southwest of Baltimore (39.2552N, 76.6753W), NDC in the Tenleytown neighborhood of
183 northwest Washington, DC (38.9499N, 77.0796W), and ARL in Arlington, VA (38.8917N,
184 77.1317W). Additionally, the NOAA / University of Virginia CO₂ observation site in
185 Shenandoah National Park (SNP, 38.6170N, 78.3500W; Lee et al., 2012; Andrews et al., 2014;
186 CarbonTracker Team, 2017) was used as a rural comparison site, as it is located at 1008 m above
187 sea level putting it frequently above the surrounding local planetary boundary layer (PBL) at
188 night (Poulida et al., 1991). This site is also over 10 km from the nearest town, over 25 km from
189 the nearest major highway, and far from most local anthropogenic influences. This analysis uses
190 observations from all inlet heights when available, but for plotting purposes only the lowest inlet

191 is shown for time series. These four site locations are shown in Fig. 1, with additional
192 information in Table 1.



193
194 **Figure 1.** Map showing the WRF-Chem domain configuration used in this analysis. Domain d01
195 is modeled with 9km horizontal resolution, d02 with 3km, and d03 with a 1km horizontal
196 resolution. The lower right inset shows the immediate area around d03 and the locations of the
197 observing sites used: Shenandoah National Park (SNP; red circle), Arlington, VA (ARL; green
198 circle), Northwest Washington, DC (NDC; yellow circle), Halethorpe, MD (HAL; blue circle),
199 and Washington National Airport for wind direction (KDCA; black square). Major highways are
200 plotted as dark gray lines on the inset map along with the county boundaries in light gray.

201
202
203
204

Site	SNP	ARL	NDC	HAL
Site Name	Shenandoah National Park	Arlington, VA	Northwest Washington, DC	Halethorpe, MD
Latitude (°N)	38.6170	38.8917	38.9499	39.2552
Longitude (°W)	78.3500	77.1317	77.0796	76.6753
Inlet Heights (meters above ground level)	17 m	50 m, 92 m	45 m, 91 m	29 m, 58 m
Site Elevation (meters above sea level)	1008 m	111 m	128 m	70 m
Data provider	NOAA/UVA	NIST/EN	NIST/EN	NIST/EN

205 **Table 1.** Summary of the four observation sites used in this study.
206

207 **2.2 WRF-Chem**

208 A triply nested grid was defined for our WRF-Chem model configuration (Fig. 1). The outermost
209 domain (d01) covers roughly the northeastern quadrant of the United States at a horizontal
210 resolution of 9 km. The d01 extent was chosen because the predominant wind direction for the
211 NEC-B/W is from the North and Northwest (Whelpdale et al., 1984) in February, and this extent
212 generally captures the incoming CO₂ from areas as far away as Chicago, IL. Within this parent
213 domain is an intermediate two-way nested domain (d02) with a resolution of 3 km. An additional
214 fine-scale domain (d03) is nested within d02; it features a horizontal resolution of 1 km that
215 covers the metropolitan areas of the NEC-B/W. A description of the parameterizations and
216 options used for each WRF-Chem domain is provided in Table S1 in the supplemental
217 information (Chou et al., 2001; Hong et al., 2004,2006; Kain 2004; Mlawer et al., 1997; Tewari
218 et al., 2004).

219
220 Meteorological initial and boundary conditions are provided by the National Oceanographic and
221 Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP)
222 North American Regional Reanalysis (NARR), a product with a horizontal resolution of 32 km,
223 30 vertical layers, and three-hourly output (Mesinger et al., 2006). Because the simulation runs
224 for the entire month, sea surface temperatures are also included as boundary conditions in the
225 WRF-Chem model. The high-resolution version of the NOAA NCEP real-time, global, sea
226 surface temperature analysis (RTG_SST_HR) with a horizontal resolution of 1/12 degree and
227 daily output is used (Thiébaux et al., 2003; Gemmil et al., 2007). No analysis nudging or data
228 assimilation is performed as part of the modeling work, as there is concern about how these
229 techniques may affect the simulated CO₂ fields when optimizing meteorology.

230
231 Initial and boundary conditions for the background CO₂ concentrations are provided by NOAA
232 Earth System Research Laboratory's (ESRL) CarbonTracker Near Real-Time gridded product

233 (Peters et al., 2007; <https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT-NRT/index.php>).
234 This is a 3D mole fraction product with three-hourly output and a horizontal resolution of 1° over
235 North America. This background value is available as a separate tracer at all hours of the
236 simulation, and is added to the other tracers for the total predicted CO₂. All tracers resulting from
237 the anthropogenic emissions inventories have initial and boundary conditions of zero ppm. The
238 model-simulated CO₂ mole fraction determined at a point in time and space is the sum of the
239 tracer associated with the specified anthropogenic emissions inventory, the biospheric flux
240 tracer, and the advected background CarbonTracker mole fraction.

241
242 For atmospheric transport and trace gas dispersion, three of the most important meteorological
243 variables are the PBL height, near-surface wind speed, and near-surface wind direction.
244 Averaged over 10 meteorological surface sites, the WRF run over-predicts wind speed by 1.2
245 m/s and has a wind direction bias of 2.8°, however each hour can significantly vary. These results
246 are consistent with previous comparisons of WRF to observations in similar work (e.g. Nehr Korn
247 et al., 2012; Feng et al., 2016). Conversely, WRF’s YSU PBL scheme tends to underpredict PBL
248 heights, with an average value of -70 m from observations at three airports over the entire month.
249 Observations of PBL height are computed from potential temperature profiles from both
250 commercial aircraft and radiosondes; both are limited in temporal and spatial resolution resulting
251 in gaps, particularly during the overnight hours. More details on the evaluation of WRF’s
252 performance relative to meteorological observations are available in the supplemental
253 information, with specific scenarios presented in Sect. 3.2.

254

255 **2.3 Emissions Inventories**

256 To evaluate whether the modeled observations are impacted more by the underlying emissions or
257 transport, an ensemble of tracers of atmospheric CO₂ resulting from different emissions
258 inventories are used within the same transport model simulation using WRF-Chem. For this
259 study, four different anthropogenic CO₂ emissions inventories are used: EDGAR, FFDAS,
260 ODIAC, and Vulcan. Since the inventories were generated for a year differing from the modeled
261 year, ratios are used to scale each emission product using national totals from the U.S. Energy
262 Information Administration’s (EIA) Monthly Energy Review
263 (<https://www.eia.gov/totalenergy/data/monthly/>) for each day of February as shown in Equation
264 1. Emission products will be referred to as tracers from henceforward to correspond to WRF-
265 Chem nomenclature. Fig. 2 shows a map of the hourly mean flux of CO₂ from each inventory
266 interpolated to all three WRF-Chem domains. The following paragraphs and Table 2 provide
267 details of each inventory.

268

$$269 \quad \text{Inventory}_{2016} = \text{Inventory}_{Feb.Year} \times \frac{EIA\ Total_{Feb.2016}}{EIA\ Total_{Feb.Year}} \quad (1)$$

270

271 The first inventory employed is EDGAR, i.e. the Emissions Dataset for Global Atmospheric
272 Research version 4.2 (Olivier et al., 2005; <http://edgar.jrc.ec.europa.eu>). EDGAR is a global
273 emissions product with a horizontal resolution of 0.1°, and provides average fluxes for the year
274 2010 based on the International Energy Agency’s (IEA) energy budget statistics (IEA, 2012).
275 The emissions are then distributed on the 0.1° x 0.1° grid by incorporating population density,
276 road networks, and the locations of point sources and industrial processes.

277

278 The Fossil Fuel Data Assimilation System (FFDAS; Rayner et al., 2010; Asefi-Najafabady et al.,
 279 2014) is also used. As with EDGAR, FFDAS is a global product with a horizontal grid of 0.1° x
 280 0.1° but unlike EDGAR, it features hourly varying anthropogenic fluxes for the entirety of 2015.
 281 FFDAS utilizes the Kaya Identity, a method to estimate emissions based off of economic factors,
 282 as well as information on national fossil fuel CO₂ emissions, satellite-derived nightlights,
 283 population density, and power plant information to estimate flux at each grid point.

284
 285 The Open-source Data Inventory for Anthropogenic CO₂ (ODIAC; Oda and Maksyutov, 2011;
 286 Oda and Maksyutov, 2015; Oda et al., 2018) is the third inventory used in WRF-Chem. It is the
 287 only dataset of the four chosen with a finer horizontal resolution of approximately 1 km, or ~0.01
 288 °. Using the total emissions estimated by the Carbon Dioxide Information and Analysis Center
 289 (CDIAC) at the US Department of Energy’s Oak Ridge National Laboratory, the locations of
 290 point sources and satellite-derived nightlights are then used to distribute the emissions onto the 1
 291 km grid. Monthly total fluxes are provided by ODIAC for each month projected using statistical
 292 data from the energy company BP with the most recent version for the year 2015.

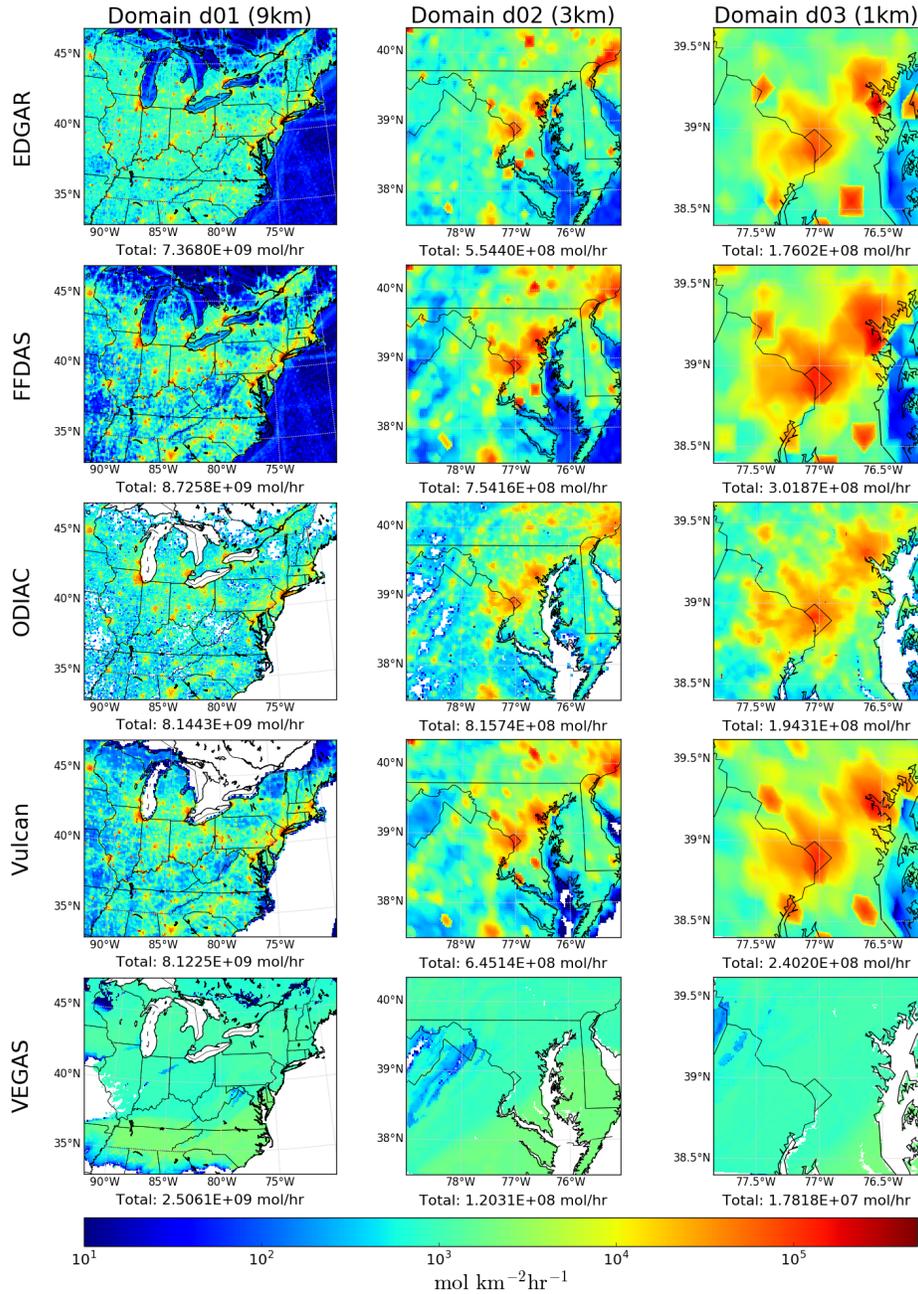
293
 294 The fourth inventory is Vulcan (Gurney et al., 2009) which is a 10 km x 10 km fossil fuel
 295 emissions dataset for the United States for the year 2002. Unlike the others, the emissions of
 296 individual buildings, power plants, roadways, and other sectors are each characterized, and then
 297 aggregated to this 10 km x10 km grid. This provides a higher level of detail both spatially and
 298 temporally, but with the limitations of being much older than the other inventories, and only
 299 covers the coterminous United States, so parts of domain d01 have no emissions in this tracer.

300

Inventory	EDGAR	FFDAS	ODIAC	Vulcan
Version	4.2	2.2	2016	2.2
Horizontal Resolution	0.1 degree	0.1 degree	0.01 degree	10 km
Created for Year	2010	2015	2015	2002
Temporal Resolution Provided	Yearly	Hourly	Monthly	Hourly
Domain d01 Average Sum (Tonnes C / hr)	88416	104709	97732 106231 * (no TIMES scaling)	97469
Domain d03 Average Sum (Tonnes C / hr)	2112	3622	2332 2534 * (no TIMES scaling)	2882
Citation	Olivier et al., 2005	Rayner et al., 2010; Asefi-Najafabady et al., 2014;	Oda and Maksyutov, 2011; Oda et al. 2018	Gurney et al., 2009
Available at:	http://edgar.jrc.ec.europa.eu	http://www.gurneylab.org/portfolio-item/ffdass/	http://db.cger.nies.go.jp/dataset/ODIAC/	http://vulcan.project.asu.edu

301
 302 **Table 2.** Summary of the four anthropogenic CO₂ emissions inventories used within this study.
 303 For ODIAC, both the domain average sum is provided with and without temporal scaling added
 304 to the dataset.

Hourly Average Carbon Flux for February 2016



305
 306 **Figure 2.** Average CO₂ hourly fluxes for the four emissions inventories and the VEGAS
 307 biospheric model for all three WRF-Chem domains in February 2016. Non-positive values (zero

308 and negative) are shown as white. The hourly average flux for the month for each dataset
309 summed over the entire domain, is shown below each map.

310
311 The inventories are interpolated in time and space to ensure consistency. Each one is linearly
312 interpolated from its native grid to the three WRF-Chem domains. Mass is mostly conserved
313 however slight differences from the original datasets may occur due to the domain not lining up
314 exactly with the lower resolution pixels. Additionally, the Temporal Improvements for Modeling
315 Emissions by Scaling (TIMES; Nassar et al., 2013) scale factors are applied to ODIAC and
316 EDGAR to provide weekly and diurnal variations to these two inventories. However, we also use
317 the native monthly ODIAC product as input to WRF-Chem (aka ODIACFIX) as it allows us to
318 investigate the impact of diurnal and weekly varying fluxes on simulated observations. Note that
319 TIMES scaling results in an approximate decrease of emissions of 8.5% when averaged over the
320 entire month because while the daily average remains the same for weekdays, the scaling factor
321 causes a reduction for weekend hours and the number of each day of the week is not the same in
322 any given month. The impact of the TIMES scaling on the simulated observations will be
323 discussed in Section 3.1. We further ensure consistency between the inventories by shifting the
324 inventories with time information so that the calendar days and hours are the same across all
325 emission products. For example, the fluxes for February 2, 2015 of FFDAS are used for
326 February 1, 2016 as they are both Mondays.

327
328
$$Inventory_{2016} = Inventory_{Feb.Year} \times \frac{EIA\ Total_{Feb.2016}}{EIA\ Total_{Feb.Year}}$$

329 As mentioned earlier, a dynamical vegetation model has also been coupled to WRF-Chem to
330 provide the contribution of biogenic fluxes to the simulated observations. The VEGAS
331 Global-Atmosphere-Soil (VEGAS) model (Zeng et al., 2005) is coupled offline with WRF-Chem
332 to provide hourly biospheric CO₂ flux. Because VEGAS features carbon pools and dynamic
333 vegetation growth, the model must first be spun up on the domain to achieve a climatology. For
334 this analysis, VEGAS is first initialized by forcing it with the Climate Forecast System
335 Reanalysis version 2 (CFSRv2; Saha et al., 2014) calibration climatologies for the years 1981 to
336 2010. The CFSR climatology dataset is first regridded to the WRF-Chem domains, and then the
337 model is run for 100 years using this calibration climatology repeatedly to reach equilibrium. To
338 generate the land to atmosphere carbon flux, VEGAS uses the WRF-Chem meteorological
339 output variables (2 m temperature, 2 m specific humidity, hourly precipitation, 10 m winds, skin
340 temperature, and total net radiation) as well as the WRF domain topography, emissivity, and
341 albedo. Figure 2 shows the hourly average biospheric flux from VEGAS on all three WRF-Chem
342 domains in February 2016; including areas of net uptake (the white region in the south part of
343 domain d01). The biospheric VEGAS tracer has been added to all the simulated CO₂ values
344 shown although its average contribution to the simulated CO₂ mole fraction across all sites for
345 the month is approximately 1 ppm, with a standard deviation of 0.97 ppm due to the diurnal
346 cycle and meteorological fluctuations. Additionally, no evaluation has been performed on the
347 biospheric fluxes from VEGAS at these spatial scales further highlighting why we focused on
348 winter months for this analysis.

349
350 **3 Results**

351 In this section we assess WRF-Chem’s ability to simulate the atmospheric CO₂ in the NEC-B/W
352 by comparing modeled CO₂ mole fractions at four locations to high-accuracy in situ observations
353 from the three urban and one rural tower sites. First, these datasets are compared over the entire

354 month-long simulation to determine the overall performance of the model. We then select
355 specific time periods of the month to diagnose possible causes of both high and low performing
356 scenarios. These two analyses help us to evaluate the performance of WRF-Chem in modeling
357 transport and dispersion of urban CO₂ and whether there is sufficient skill in the model for use
358 with various GHG flux estimation methods.

359

360 **3.1 Overall Model Performance**

361 **3.1.1 Tower Observations**

362 Not surprisingly, the magnitudes and variability of the observations from the towers are different
363 for the rural site compared to those from the urban towers (Fig. 3). Over the four locations, the
364 lowest observed CO₂ of the four sites was typically at SNP (with an average of 412.2 ppm vs. the
365 mean of the urban sites at just over 421 ppm) due to its rural location and high altitude,
366 frequently above the PBL in the free troposphere. The amplitude of the observational diurnal
367 cycle at SNP is also smaller than the urban towers but can vary from day to day depending on the
368 synoptic weather situation. The variability in the diurnal cycle at the urban sites is much greater
369 (frequently as high as 50 ppm, but occasionally under 5 ppm) which indicates that synoptic
370 events have a large impact on urban CO₂ observations given the magnitude and variation of the
371 underlying flux distribution in such areas.

372

373 **3.1.2 Simulated Observations**

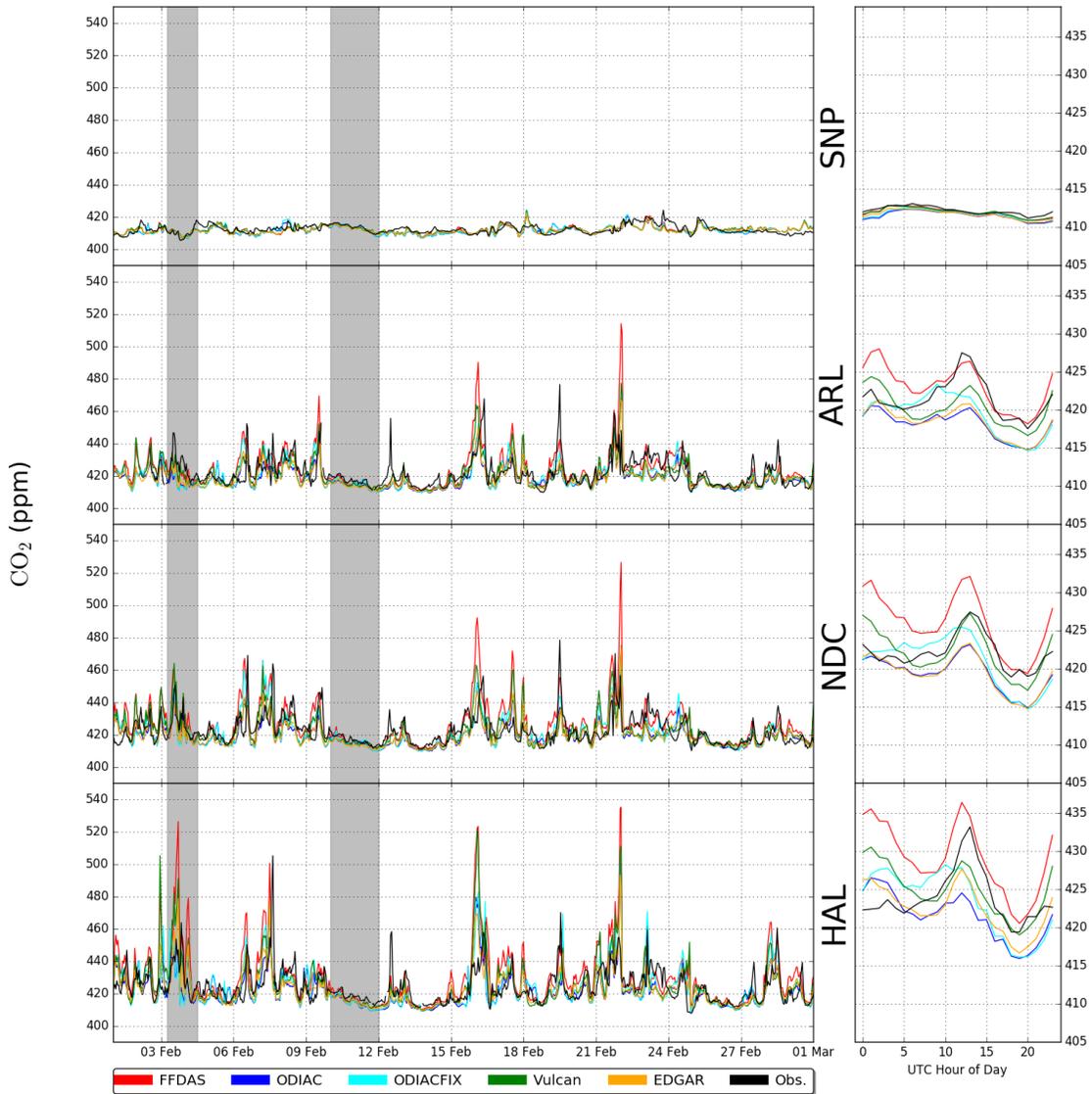
374 In general, WRF-Chem generated mole fractions with similar magnitudes and variabilities to the
375 observed mole fractions from the four tower sites. Consistent with the observations, the relative
376 magnitudes of the simulated mole fractions at urban towers are larger than those from the rural
377 site and they exhibit more diurnal variability (the right panels of Fig. 3). In addition, when
378 looking at the model predicted mole fractions in Fig. 3 across all sites, one can note the
379 variations in synoptic and diurnal cycles are similar to the observed time series. These results
380 provide evidence that the model is able to reasonably recreate the time series of CO₂ mole
381 fractions when looking at the complete time series, but an in-depth analysis is required to
382 determine its performance for a particular day or period.

383

384 During certain unfavorable meteorological conditions, the spread of the individual emission
385 tracers increases even though their overall variability remains proportionally the same. In terms
386 of overall magnitudes, the differences in the minimum and maximum daily values (i.e.
387 differences between the observed and modeled mole fractions) can be at times quite significant.
388 These large differences correspond to synoptic scale weather patterns (3 d to 5 d) that also create
389 the variability in the observations as discussed in Section 3.1.1. Depending on the day, the
390 differences between modeled and observed CO₂ mole fractions throughout the day can vary by
391 an order of magnitude from less than 5 ppm to over 50 ppm. These synoptic weather conditions
392 will be discussed further in Section 3.2.

393

Modeled and Observed CO₂ Timeseries in February 2016 and Diurnal Average



394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408

Figure 3. Time series of hourly averaged modeled versus observed CO₂ mole fractions at four observing sites for all hours of the day on the left and diurnal hourly averages on the right. The black lines are the observed values, and each color represents the model-simulated CO₂ interpolated to that location and inlet height (only the lowest inlet levels are plotted at the 3 urban sites). The model-simulated mole fraction at a point in time and space is the sum of an anthropogenic tracer generated from a specific inventory plus the VEGAS biospheric flux tracer plus the background CarbonTracker advected value. From top to bottom: Shenandoah National Park (SNP), Arlington, VA (ARL), Northwest Washington, DC (NDC), and Halethorpe, MD (HAL). FFDAS predicted values are in red, ODIAC in blue, ODIAC without temporal scaling in light blue, Vulcan in green, and EDGAR in orange. Gray shaded areas are scenarios described in detail in Section 3.2.

409 To characterize the performance of the simulated mole fractions using WRF-Chem relative to
410 observations while accounting for the differences in emissions inventories, a linear regression
411 was performed for all five modeled time series at each site (and both inlets where applicable)
412 against their respective observed time series. The results from these regressions are available in
413 the supplemental information Table S2. For this regression, outliers, defined as differences larger
414 than three standard deviations between the mean of both the observations and the WRF predicted
415 values are removed. The resulting statistics help discern whether the simulated mole fractions, on
416 average, deviate strongly from the observations, are linearly related, and reflect the true
417 variability.

418
419 Overall, for all of the urban locations, FFDAS has the strongest linear relationship to the
420 observations with a slope closest to one. However, FFDAS has the largest R^2 for only four out
421 of the seven timeseries (those from two inlet heights at each of the three urban towers and from
422 SNP) for the February 2016 model simulation. However, the lowest R^2 value is associated with
423 the modeled mole fractions using FFDAS at both the HAL inlet levels. This could potentially be
424 due to the location of HAL near large FFDAS point sources in Baltimore that are redistributed
425 onto the native WRF-Chem grid. For the other inventories, the slopes and R^2 varies across all
426 towers.

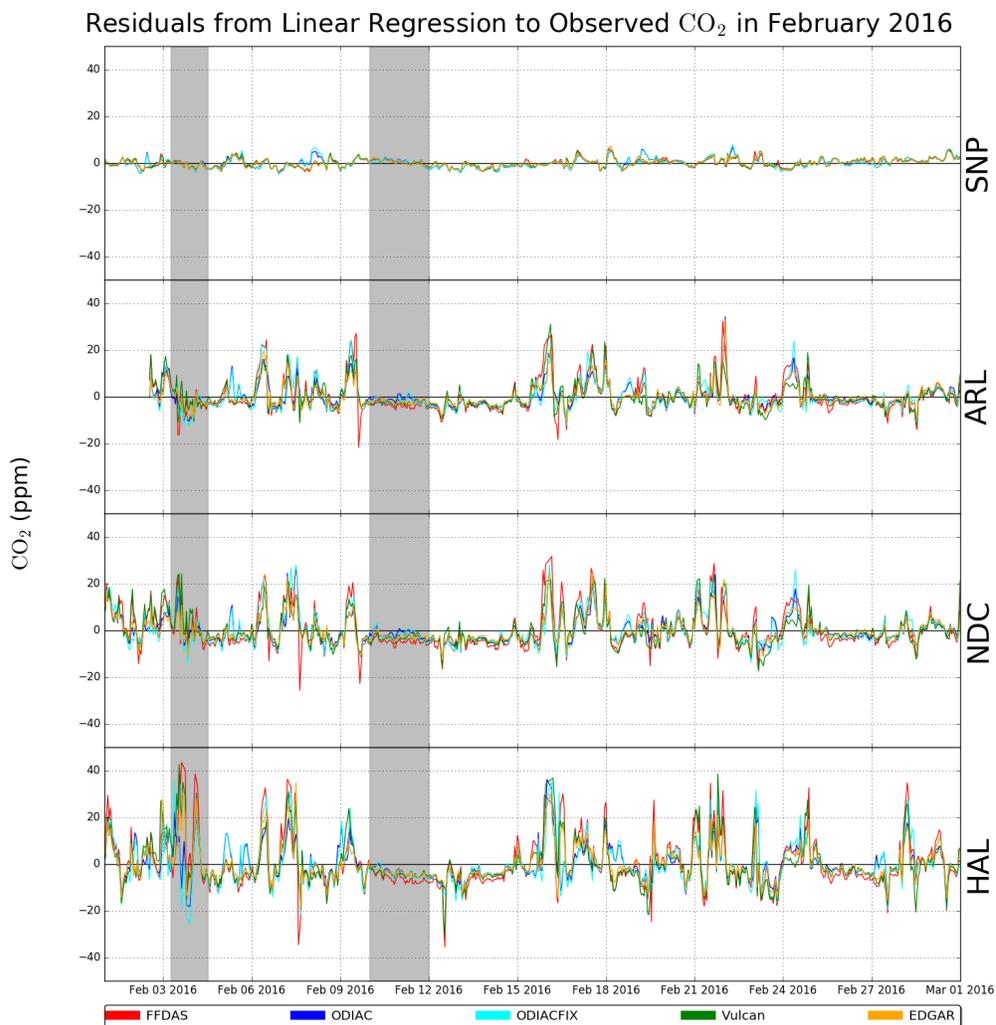
427
428 We also performed a two-day running correlation on each modeled CO_2 timeseries against the
429 observations and a plot for each of the four observing sites is available in the supplemental
430 information in Fig. S2. Similar to what is shown in the residuals of the linear regression, there
431 are times when the modeled CO_2 mole fraction is highly correlated with the observations and
432 periods where correlation is low, and even occasionally negatively correlated. All four observing
433 sites have similar ranges, with two-day correlations ranging from as low as approximately -0.5 to
434 over 0.98. The curves across the four observing sites all look very similar, suggesting that the
435 correlations are consistent across the domain and not limited to one location.

436
437 The slopes associated with simulated observations from the other four anthropogenic tracers
438 have a small range of spread between inventories, with the spread being between 0.03 and 0.16
439 depending on observing site. The slopes closest to zero tend to be either ODIAC or EDGAR
440 depending on the observing point location. The fact that these two inventories have similar
441 slopes is not totally unexpected, as their derived emissions may be distributed spatially in a
442 similar manner, albeit at different resolutions, and the TIMES scaling factors are applied to both
443 of them.

444
445 The other statistics indicate that the performance of the modeled time-series is dependent on
446 local conditions (i.e. meteorology or flux distribution) around each tower site since no single
447 tracer consistently out-performs the others. For example, the mean absolute error (MAE)
448 associated with the five anthropogenic tracers varies across tower inlets for a single inventory
449 (such as FFDAS from 3.97 ppm to 7.56 ppm or Vulcan from 4.40 ppm to 6.73 ppm). FFDAS
450 generally has the highest MAE with EDGAR or Vulcan usually having the lowest, but again,
451 there is no consensus on any best or worst performer at each observing site. This suggests that
452 model performance should not be assessed on average across the entire domain but rather locally
453 around tower sites.

454

455 After calculating these linear regressions, the fitted datasets can be subtracted from the original
 456 modeled time series to see where the linear fit is not valid. Figure 4 shows the residuals of each
 457 linear fit from the observed CO₂ at each site where the five colors represent the different
 458 anthropogenic tracers in WRF-Chem. As with the slope, FFDAS (red) has a consistently larger
 459 absolute residual value (5.1 ppm) than the other four datasets (ODIAC (blue): 3.6 ppm;
 460 ODIACFIX (cyan): 4.3 ppm; Vulcan (green): 4.3 ppm; EDGAR (orange) 3.5 ppm) for February
 461 2016 across the observing sites, likely due to the periodic high values skewing the linear fit as
 462 noted earlier. The residual plot also shows clearly periods where the simulated CO₂ deviates
 463 greatly from the observations for all tracers. This suggests that at times 1) the synoptic scale
 464 background CO₂ provided by CarbonTracker may not be resolved correctly, 2) there are
 465 sufficient errors in the meteorological transport, or 3) VEGAS is under-predicting respiration
 466 during this period. Two of these cases (the dark gray shaded regions in Fig. 4) will be described
 467 in detail in Section 3.2.
 468



469
 470
 471 **Figure 4.** Residuals of a linear regression between the observed CO₂ and each tracer at all four
 472 observing sites (lowest inlet only at the 3 urban sites). See Section SI2 of the supplemental
 473 information for the regression equation used for this analysis. The different colors represent the

474 five different tracers from the multiple emissions inputs. The dark gray shaded areas are
475 scenarios described in detail in Section 3.2 which were also shown on Figure 3.

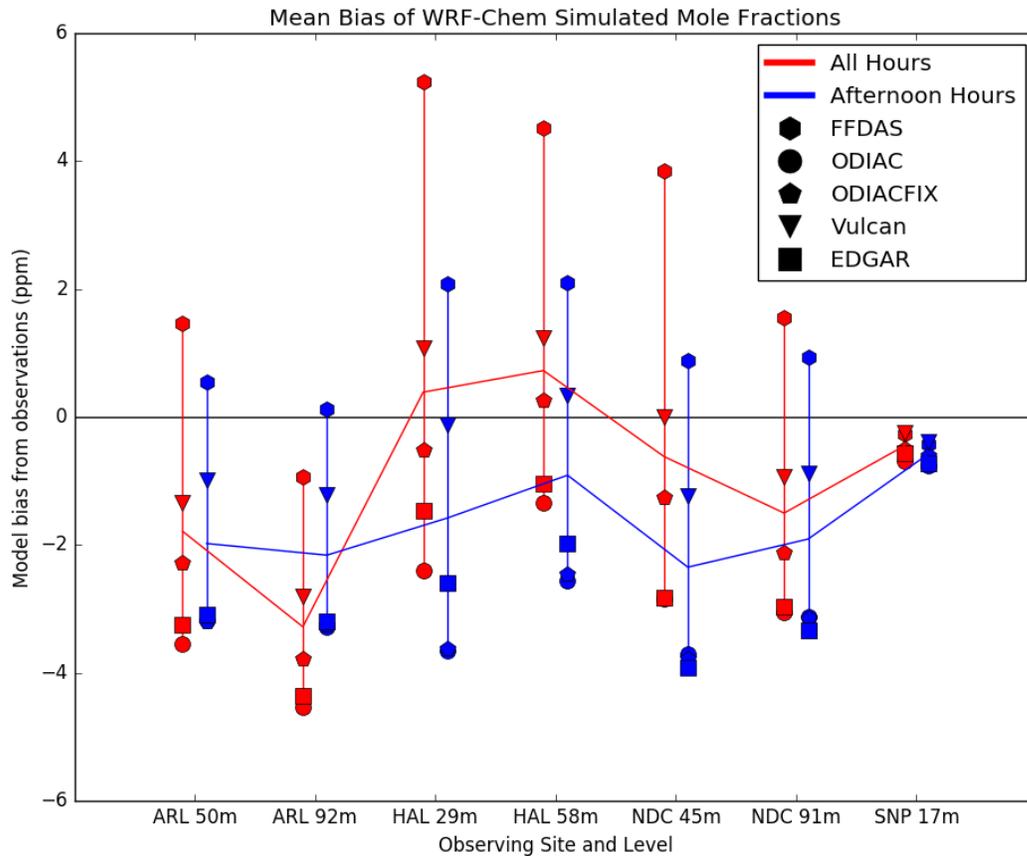
476
477 To investigate the impact of bias on the modeled vertical mixing and its representation of the
478 PBL, the simulated mole fractions are analyzed using (1) all hours of the time series, and (2)
479 afternoon hours only (12 pm to 4 pm local time; 17 UTC to 21 UTC), both without removing any
480 outliers. Afternoon observations are typically used in inversions since it is assumed that the
481 meteorological models can better represent well mixed conditions (i.e. the lower relative error of
482 the higher PBL heights) typically found in the middle of the day (e.g., McKain et al., 2012; Kort
483 et al., 2013; Breón et al., 2015; McKain et al., 2015; Lauvaux et al., 2016; Sargent et al., 2018).
484 Figure 5 shows the monthly mean bias of simulated CO₂ mole fractions for all hours and
485 afternoon hours only for all five tracers at the observing sites and different inlet heights.

486
487 The spread of the monthly bias from all tracers ranges from -4.5 ppm to 5.2 ppm for the five
488 simulated observational timeseries encompassing all hours of the day. FFDAS has a slight
489 positive mean bias at all three urban sites (the largest being at HAL and at NDC for the lowest
490 inlet). This is consistent with the domain mean hourly averaged flux in domain d03 being the
491 largest for FFDAS as shown in Fig. 2 and Table 2. Conversely, as expected from the flux
492 summaries in Section 2.3, the most negative bias tends to be from either ODIAC or EDGAR as
493 they are diurnally scaled using the TIMES dataset and have the lowest domain mean hourly
494 averaged flux in the urban domain. The impact of the TIMES scaling is clearly demonstrated by
495 the fact that ODIACFIX has a smaller bias than that of ODIAC.

496
497 When considering only afternoon hours, the spread in the monthly bias is smaller, ranging from
498 (-3.9 ppm to 2.1 ppm). The FFDAS tracer yields the highest simulated CO₂ mole fraction
499 (similar to all hours), and is the only inventory that has a clear positive bias. Although mostly
500 negative, the Vulcan tracer has a near zero bias at both the HAL inlets. When looking at
501 afternoon hours only, the ODIAC and ODIACFIX tracers are virtually the same as the TIMES
502 scaling factors are based on a value of one for the mid-afternoon (the emissions are scaled down
503 overnight). The mean bias from EDGAR during these periods is similar to that from ODIAC,
504 with a negative mean bias of approximately 3 ppm to 4 ppm from the observations. This range of
505 biases (FFDAS positive to ODIAC and EDGAR being the lowest) is consistent with their
506 respective rankings in the areal total anthropogenic flux for February 2016. Using afternoon
507 hours only appears to reduce the spread of the modeled time series overall compared to the
508 complete datasets but this is dominated by the large reduction in spread at HAL and the lower
509 inlet at NDC.

510
511 Generally, WRF-Chem using these emissions inventories tends to underpredict near-surface CO₂
512 as shown in Fig. 5 (red and blue lines represent the average associated with all hours and
513 afternoon hours respectively). On average, the mean of the five modeled time series is below the
514 observed values, except for the all-hours datasets at HAL, likely largely the result of an
515 underestimation of emissions. The smaller bias at SNP implies that a bias in the CarbonTracker
516 background or the biospheric flux may also contribute to the low bias across the domain.
517 Virtually no spread (< 1 ppm) is found at SNP for both all hours and afternoon only, with very
518 little changes between the two periods, consistent with its rural location and lack of enhanced
519 CO₂ values from emissions sources. This general underprediction trend is exacerbated when only

520 including afternoon hours (the blue line in Fig. 5), where the mean bias becomes more negative
 521 for all datasets except for the higher inlet level at ARL. However, when including all hours (the
 522 red line in Fig. 5), the absolute value of the average mean bias across all inventories is smaller
 523 than when only looking at afternoon hours. This is likely due to the overprediction that
 524 commonly occurs at night balancing out the underprediction occurring in the afternoon hours.
 525 ~~Because of this, atmospheric inversion analyses may want to use observations from all hours~~
 526 ~~rather than just afternoon to better constrain the estimates.~~
 527



528 **Figure 5.** Mean bias of WRF-Chem simulated CO₂ mole fractions (ppm) compared to
 529 observations at each observing site and for both inlets where applicable for all five tracers
 530 (FFDAS: square, ODIAC: circle, ODIACFIX: pentagon, Vulcan: triangle, and EDGAR: square)
 531 during all hours (red) and during afternoon (12 pm to 4 pm local time; blue) hours only. Means
 532 of bias at each inlet are connected with additional lines: red for all hours and blue for afternoon.
 533
 534

535 3.2 Typical Scenarios

536 The model-data comparisons presented in Section 3.1 cover the entire month of February 2016.
 537 While the mean biases of the modeled CO₂ mole fractions are relatively small, the variation
 538 between days, and even between individual hours of the day, can be significant, as shown in the
 539 residual plots in Fig. 4 (as well as in the running correlation plots in Fig. S2). Two sample cases
 540 (the dark gray shaded regions in Fig. 4) are presented to show a range of scenarios. One of these

541 cases is when the model error is large (greater than 10 % of the observed total value) and the
542 proportion of variability (the ratio of the standard deviation of the differences over the period of
543 interest and the standard deviation over the entire month) is greater than 100 %. The second
544 scenario occurs when the model-observation difference is small ($\sim < 1$ % of the total CO_2) and the
545 proportion of variability relative to the whole month is under 40 %. For trace gases in the
546 atmosphere, winds (both speed and direction) as well as the height of the PBL are the most
547 important meteorological factors in estimating near-surface CO_2 . In the subsequent subsections,
548 two scenarios are presented that show examples of how different synoptic weather situations can
549 affect the ability of WRF to predict these variables, and in turn, can impact the quality of the
550 predicted CO_2 in a forward transport model.

551

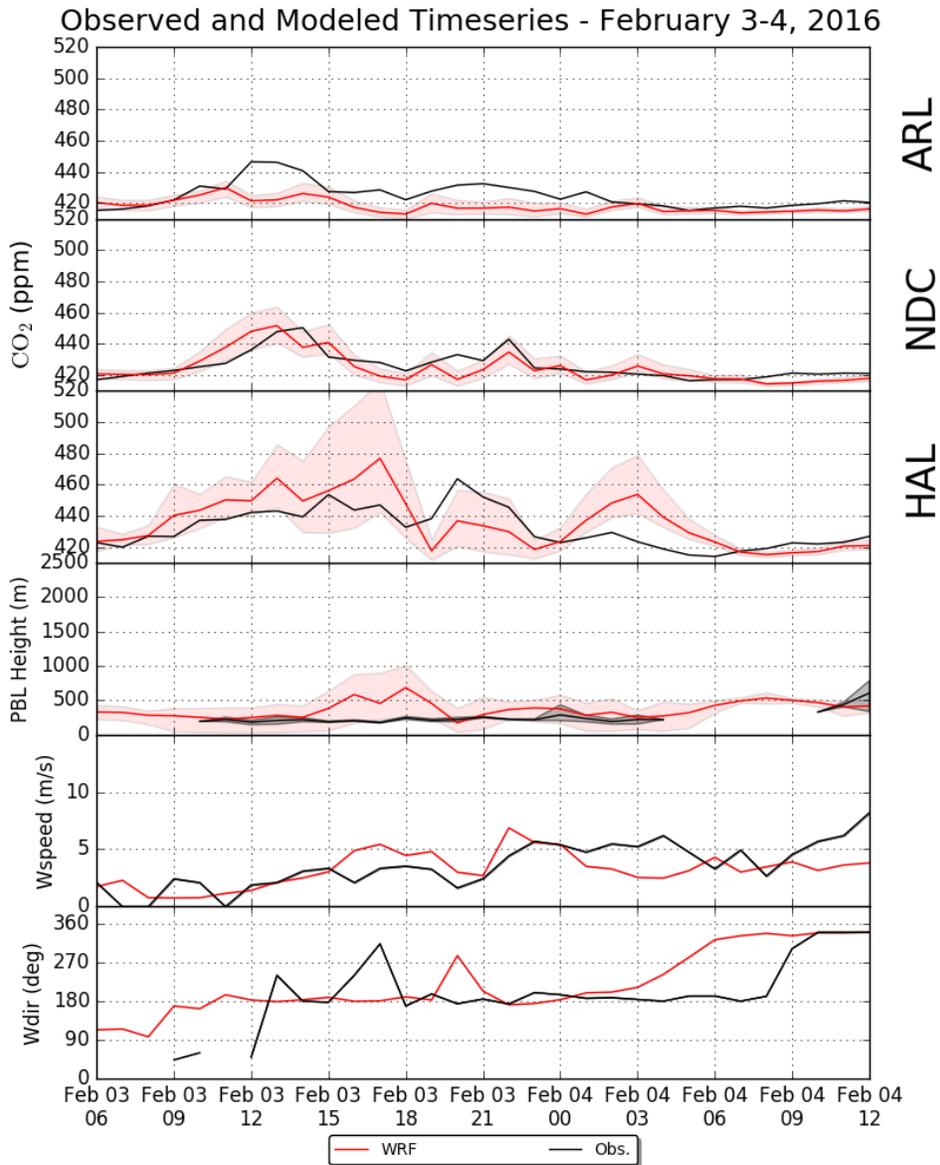
552 **3.2.1 Frontal Passages**

553 A typical mid-latitude location such as the NEC-B/W experiences frontal passages every 3 to 7
554 days. When a front passes over an area, such as this urban domain, it causes sharp changes in
555 wind speed and direction, temperature and moisture content, as well as other defining features of
556 an air mass including CO_2 mole fractions (Parazoo et al., 2008). Figure 6 shows an example of
557 the impact on CO_2 mole fractions when two fronts pass over the NEC-B/W as a mid-latitude
558 cyclone moves east. The gradual slope of the warm front (~ 12 UTC February 3) can create a
559 shallow PBL, allowing CO_2 to accumulate near the surface, whereas a cold front (~ 8 UTC
560 February 4) is much steeper in its vertical structure. For the latter, the observed wind shifts are
561 much more abrupt both in speed and direction. Both frontal passages can also create surface
562 convergence ahead of the front, allowing CO_2 to build up in the PBL immediately before the
563 wind shift occurs. Figure 7 shows surface CO_2 concentrations predicted using the FFDAS
564 emissions inventory and 10 m wind vectors during the model simulated frontal passage (5 UTC
565 February 4), illustrating the spatial gradient of CO_2 during this period.

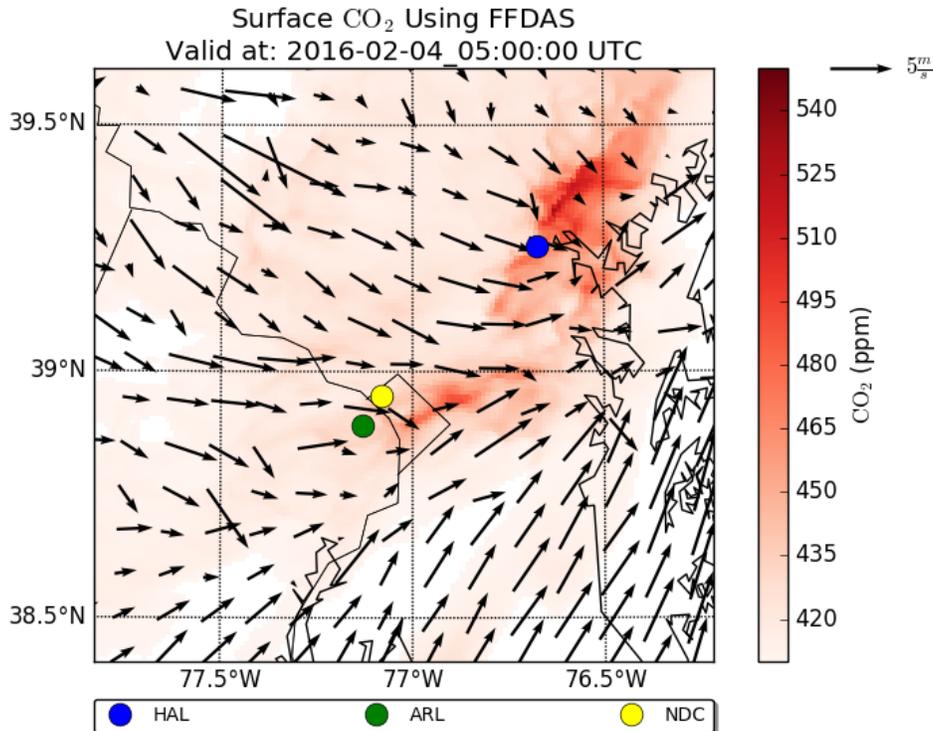
566

567 For both the warm and cold fronts, WRF simulates the frontal passage and associated wind shift
568 4 to 5 hours before the observed passage (Fig. 6). Generally, correlations are observed between
569 the wind direction, the PBL height, and the simulated CO_2 . The simulated warm front caps the
570 PBL, allowing for predicted mole fractions to increase across the domain. The diurnal range is
571 well resolved by WRF-Chem at NDC, but it is greatly overestimated at HAL and underestimated
572 at ARL, likely related to the timing and position of the passing front. The spread between the
573 various inventories at HAL is also much larger than at the other two sites. During the first frontal
574 passage (12 UTC February 3), it is clear that the predicted CO_2 values are influenced by the
575 shallow PBL depth, which is also observed at a number of periods during the month where the
576 model and observations diverge. However, during the simulated cold front (~ 8 UTC February 4),
577 the PBL heights do not change significantly, but the predicted CO_2 peaks and then drops rapidly
578 as the wind shifts from southerly to northerly. This feature is seen in the simulated time series at
579 all three sites with WRF-Chem results underestimating CO_2 mole fractions at the observing
580 locations when the front actually passes through. This example illustrates that the meteorological
581 error, in both the timing of the front and the PBL depth, dominates the error in the simulated CO_2
582 values rather than the underlying emissions, as often the spread in the emissions inventories
583 (shaded red area) is generally smaller than the difference between the model mean (red line) and
584 the observations (black).

585



586
587 **Figure 6.** Observed (black) and modeled (red line average; shaded red spread of the five
588 emissions inventories) hourly averaged CO₂ at all three urban sites for a typical frontal passage
589 period (February 3-4, 2016). The fourth panel shows the mean (black line) and spread (shaded
590 gray) of the inferred PBL heights from aircraft profiles and radiosonde observations, with the
591 YSU PBL scheme predicted PBL height from WRF at each observing site mean (red line) and
592 spread (shaded red). Bottom two panels show 10 m observed (black) and modeled (red) wind
593 speed and direction at KDCA. See the supplemental information for details on the
594 meteorological observations used.
595



596
597
598
599
600
601

Figure 7. Simulated surface CO₂ concentrations using FFDAS emissions and 10 m wind vectors during a cold front passage at 5 UTC February 4, 2016 in domain d03. Locations of the three urban observing sites are shown (HAL in blue; ARL in green; NDC in yellow).

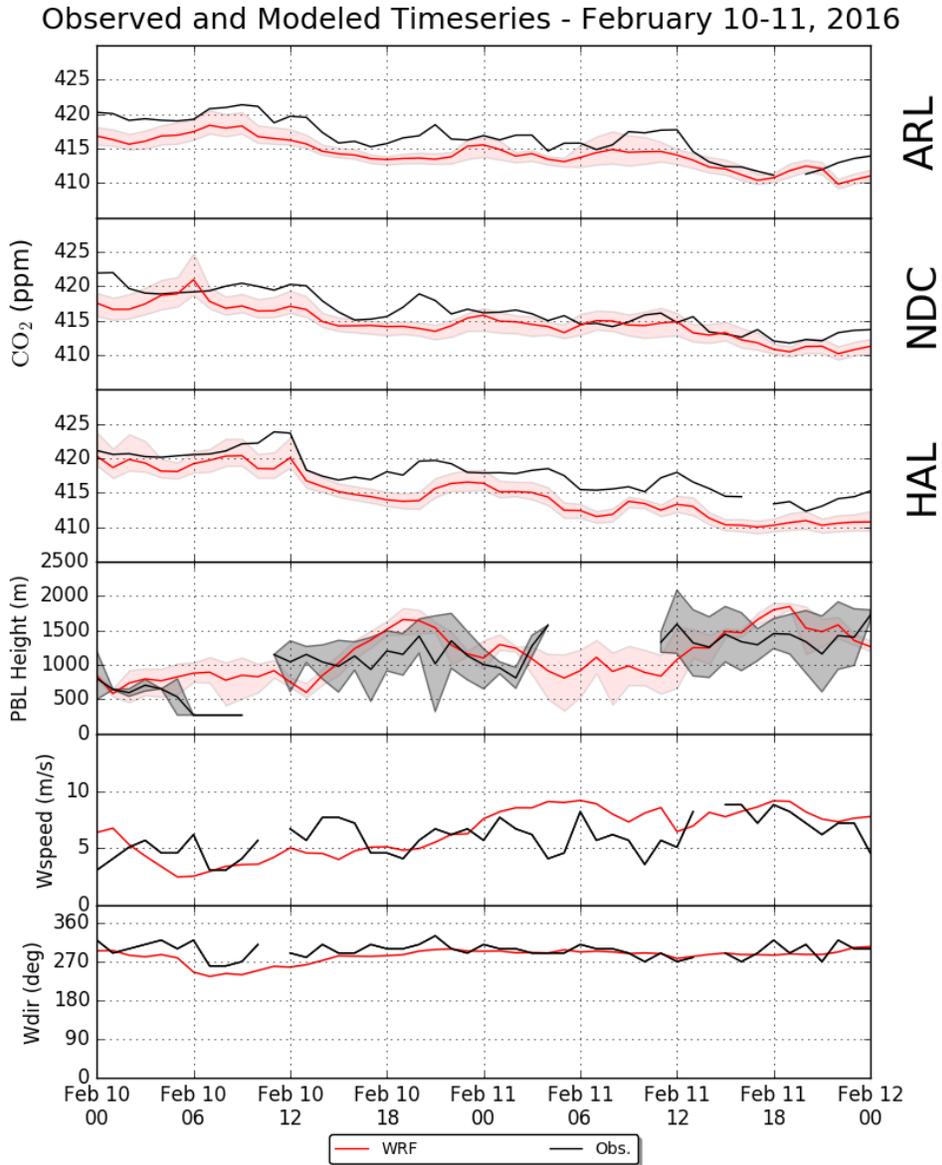
602 3.2.2 Persistent Winds

603 While the first scenario presents an example period where WRF-Chem significantly under or
604 overpredicts CO₂ compared to the observations, there are times when the model simulated CO₂ is
605 within 1 % of the observed value. One such example is from February 10 to February 11, 2016,
606 when winds are steady and from a direction where the upwind CO₂ mole fractions are more
607 representative of the global average. Figure 8 shows the modeled and observed CO₂ as well as
608 wind direction and PBL height for this period. During these two days, the wind is persistently
609 from the west or northwest, bringing in a steady stream of air into the urban area where the mole
610 fraction observations are strongly influenced by the incoming atmosphere, or regional
611 background, values.

612

613 The average modeled CO₂ mole fractions at all three sites for this 48 hour subset are always
614 within 5 ppm of the observed value. At ARL and HAL most hours are underpredicted, possibly
615 due to the overprediction of the PBL height during this period as the modeled wind speed and
616 direction agree well with the observations, but at NDC the average modeled value differs by 1
617 ppm to 2 ppm from the observations from 0 UTC to 18 UTC on February 11. Additionally, the
618 predicted CO₂ mole fractions from all five inventories do not vary significantly from one another
619 during this case, with a spread of only 2 ppm to 3 ppm on either side of the mean throughout the
620 period. Despite this agreement, there are still some subtleties that can be observed in the data. All
621 three sites have a local maximum in the observations at around 20 UTC on February 10, but the
622 modeled time series have a local minimum there. At the same time, WRF overpredicts the PBL

623 height compared to the observed height, likely causing dilution in the predicted CO₂ mole
 624 fractions. These features are present in all the inventory tracers, and often the observations are
 625 not within the spread of the five models, even during this period of relatively good model
 626 performance. In addition to the meteorological errors described above, biases in the
 627 CarbonTracker background or the biospheric tracer (both common to all five tracers) could also
 628 contribute to the overall error, but are likely limited to the magnitude of the bias observed at
 629 SNP.
 630



631
 632 **Figure 8.** Observed (black) and modeled (red line average; shaded red spread of the five
 633 emissions inventories) hourly averaged CO₂ mole fractions at all three urban sites for a typical
 634

635 period with persistent winds from a rural area (February 10-11, 2016). The fourth panel shows
636 the mean (black line) and spread (shaded gray) of the inferred PBL heights from aircraft profiles
637 and radiosonde observations, with the YSU PBL scheme predicted PBL height from WRF at
638 each observing site mean (red line) and spread (shaded red). Bottom two panels shows 10 m
639 observed (black) and modeled (red) wind speed and direction at KDCA. These figures indicate
640 that WRF-Chem is able to resolve both the wind direction and height of the PBL with reasonable
641 skill, although deviations do occur. See the supplemental information for details on the
642 meteorological observations used.

643

644 **4 Discussion**

645 As discussed in Section 3.1, WRF-Chem tends to underestimate hourly-averaged values of near-
646 surface CO₂ mole fractions when compared to observations averaged over the entire month,
647 however this is not necessarily the case when using FFDAS as shown in Fig. 5. The daily
648 maxima in simulated mole fractions from certain emissions inventories are often high relative to
649 observations, particularly when they occur during the overnight hours when the modeled PBL
650 depth is under predicted. However, the larger range of predicted values during these periods as
651 well as the overall underprediction during the afternoon hours result in lower averaged values
652 from the model than from observed values when including all sites, hours, and inventories as
653 shown in Fig. 5. Previous work comparing simulated CO₂ to observed time series in an urban
654 region also found an overall low bias, with predicted levels over certain hours/days exceeding
655 observed levels (Feng et al., 2016). There are some synoptic situations, e.g., February 10 and
656 February 11 (Section 3.2.2), where persistent winds allow for minimal errors in predicted CO₂
657 across the domain over an entire diurnal cycle, not just in the afternoon hours. During the
658 afternoon of February 10, in fact, WRF overpredicts the PBL height, and thus underpredicts the
659 near-surface CO₂. On this day, the modeled CO₂ may be more representative of reality during the
660 overnight hours than it is in the afternoon. Conversely, other days, such as February 3 (Section
661 3.2.1), with a passing mid-latitude cyclone and its associated fronts, yield much different results.
662 During this case, because of the predicted wind shift timing and the magnitude of the PBL height
663 varying from observations, WRF tends to either overestimate or underestimate near-surface CO₂
664 depending on the hour.

665

666 In addition to the synoptic meteorology, variations in emissions inventories are also reflected in
667 the predicted CO₂ mole fractions as shown in the mean biases described in Fig. 5. For the
668 outermost domain (d01), the areal sum of the hourly averaged emissions is similar in magnitude
669 (all within 10 % of the mean of the 4 inventories) (Fig. 2; Table 2). However, in the innermost
670 domain (d03), the areal sum of FFDAS is over 36 % higher than the mean and the lowest
671 inventory EDGAR is 20% below the mean of the inventories. Even though on the national scale
672 each inventory is similar, there can be substantial differences between them due to the emission
673 disaggregation methods (e.g., Hutchins et al. 2016; Oda et al. 2018) when considering mesoscale
674 modeling of CO₂, a problem being studied further in other works (e.g., Fischer et al., 2017). The
675 differences in totals and local sources could also be attributed to differing methodologies and
676 datasets included in each emissions inventory, including the exact location of point sources and
677 grid cell locations, among other things. These differences are generally reflected in the simulated
678 CO₂ levels, with FFDAS being the highest averaged over the entire month, and EDGAR the
679 lowest.

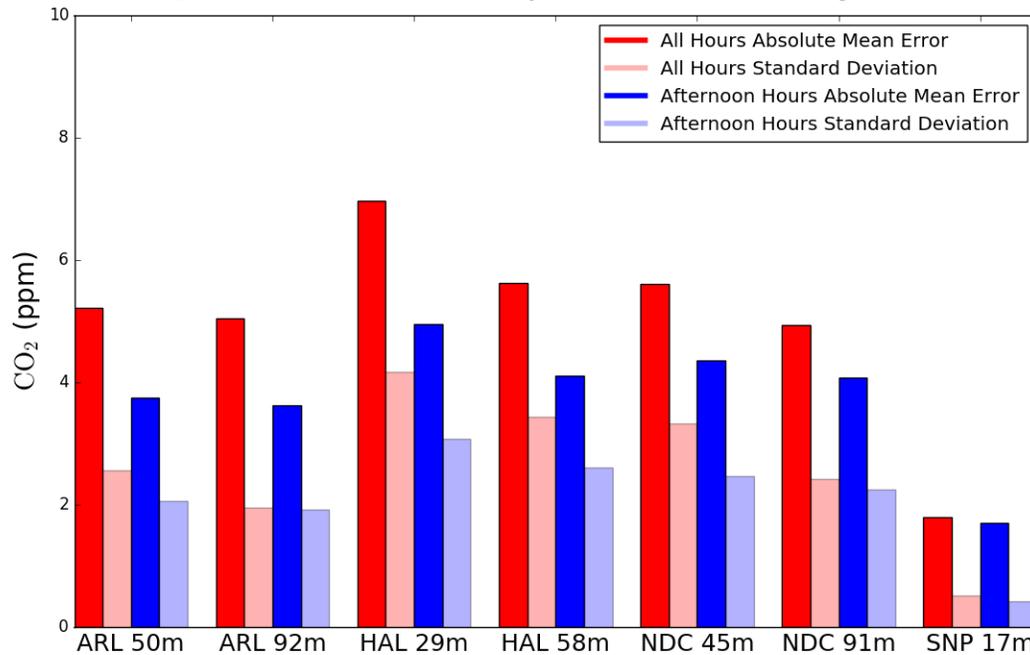
680

681 Traditionally, atmospheric inversions utilize data and meteorological model output from
682 afternoon hours (12 pm to 4 pm local time) only (Kort et al., 2013; Breón et al., 2015; Lauvaux
683 et al., 2016; McKain et al., 2015; Sargent et al., 2018). This is because the PBL is generally
684 considered to be well-mixed during this time with stronger winds and higher PBL heights,
685 allowing the relative error of these modeled quantities to be smaller during the day than during
686 the overnight hours. By including prior emissions inventories, these inversions are used to
687 estimate the total flux of carbon from an area of interest. However, in the forward modeling
688 framework presented here with WRF-Chem, using afternoon hours only may not be the best
689 metric to determine whether the model meteorology is sufficient to accurately predict CO₂ mole
690 fractions. Figure 9 shows the mean absolute error (the dark bars) as well as the mean standard
691 deviation (the lightly colored bars) of the five predicted CO₂ mole fractions for each time series,
692 both for all hours (red) and afternoon hours only (blue). When considering only the three urban
693 sites, the mean absolute error for all hours is between 3.6 ppm and 7.0 ppm across the different
694 observing sites and inlets whereas the mean standard deviations are between 1.9 ppm and 4.1
695 ppm. When including afternoon hours only, the mean absolute error of the model does decrease
696 by an average of 1.42 ppm, and the standard deviations decrease by an average of 0.58 ppm.
697 Despite the improvement in MAE when only including afternoon hours, the mean error of the
698 model is still approximately twice the variation in the predicted CO₂ values from each emissions
699 inventory (as shown in the difference between MAE and standard deviation in Fig. 9). This result
700 indicates that although limiting inversion analysis to afternoon hours may reduce overall
701 meteorological model error it also limits analysis to time periods when local and regional
702 emissions influence the observations the least and the differences between modeled mole
703 fractions from various emissions inventories are smallest (due to deeper PBLs and stronger
704 mixing).

705
706 Additionally, the mean absolute error is roughly a factor of two larger than the mean standard
707 deviation at the urban sites, with the ratio of the two ranging from 1.64 to 2.58 for all hours, and
708 1.57 to 1.89 for afternoon hours only, depending on the site and inlet level. This result suggests
709 that on average, factors common to all five tracers (meteorological error, background error, or
710 error in the biosphere tracer) contribute more to the overall model performance than the choice
711 of anthropogenic emissions inventory. Given the low bias (~2 ppm) at SNP, the extent of our
712 largest WRF domain, and small relative contribution of the VEGAS tracer to the monthly mean
713 values, we expect that the errors shown in Fig. 9 are dominated by meteorological conditions
714 during winter. This conclusion is further supported by the results of prior studies (Kretschmer et
715 al., 2014; Locatelli et al., 2015) as well as the two contrasting examples illustrated in Section 3.2.
716 However, to fully validate this conclusion, an experiment would need to be performed in contrast
717 to what is presented here, where the emissions inventory chosen is held constant and
718 meteorological transport is varied in an ensemble of simulations. It is also important to note that
719 these meteorological conditions or errors can exacerbate the differences in the emissions
720 inventories as well, like shown in Fig. 6.

721
722

CO₂ Tracer Bias and Spread in February 2016



723
 724 **Figure 9.** The mean absolute error (the dark bars) and the mean standard deviation (the lightly
 725 colored bars) of the five predicted CO₂ mole fractions for each observing site and inlet height
 726 (where applicable), both for all hours (red) and afternoon hours only (blue).
 727

728
 729 Rather than looking at the mean absolute error for the entire month and the mean standard
 730 deviation, it is also useful to consider the mean absolute error of the five tracers compared to the
 731 observed mole fraction at each hour and the associated standard deviation to get an idea of how
 732 the error and spread of the modeled CO₂ are related. These two values are reasonably related
 733 with R² values of approximately 0.3 across each of the three urban sites over the month. At
 734 times, generally when the modeled CO₂ is at its highest, approximately a factor of two difference
 735 between the highest and lowest modeled enhancements can be found. This is also when transport
 736 model errors tend to be largest such as during the overnight hours or frontal passages. But for
 737 many other cases, and on average as described above, the differences between the various
 738 emissions inventory tracers are smaller than the absolute error relative to observations. During
 739 periods of low modeled CO₂ error (i.e. small differences from the observations), the variation
 740 among the different emissions models is small but still discernable (as shown in Fig. 8 for
 741 example). Regardless, the error in modeled CO₂ compared to observations for any given hour or
 742 day appears to be influenced more by common factors such as meteorological error than the
 743 differences among the various emissions inventories. This suggests that atmospheric inversions
 744 need to attempt to quantify errors in modeled transport and dispersion as well as the uncertainty
 745 in the surface fluxes. It also suggests the need for methods that identify time periods when
 746 meteorological conditions are best represented by simulations (such as the case presented in Sect.
 747 3.2.2 compared to that in Sect. 3.2.1), rather than only including afternoon hours, and when it
 748 may be anticipated that inversions will perform with higher fidelity to actual emissions
 749 conditions. The meteorological skill could potentially be improved by using analysis nudging or

750 data assimilation techniques, but for a forward transport modeling study such as this, there are
751 some potential difficulties that could arise from assimilating both meteorological variables and
752 chemical constituents (Bocquet et al., 2015). This is particularly important for urban applications
753 where the impact of synoptic variability may strongly and adversely impact inversion analyses.
754

755 **5 Conclusions**

756 An evaluation of WRF-Chem simulated CO₂ mole fractions using multiple anthropogenic CO₂
757 emissions inventories at four CO₂ observing sites in the Baltimore, MD and Washington, DC
758 metropolitan areas was presented above. For all emissions inventories the modeled CO₂ is within
759 5 ppm of observations when averaged over all observing sites for the month of February in 2016.
760 However, for any given hour, at any particular site, the differences between the ensemble of
761 simulated CO₂ values and the observed CO₂ can vary from near zero to as high as 100 ppm (the
762 left panels of Fig. 3). The differences between the simulated time series for the different
763 emissions inventories vary significantly in time, but tend to be proportional to the magnitude of
764 the enhancement over the background CO₂ value. When averaged over the entire month all
765 simulated CO₂ mole fractions are within 8 ppm of each other (Fig. 5) representing a range of
766 approximately 2% of the total mole fraction.
767

768 This analysis suggests that the predicted mole fraction error relative to observations is dominated
769 by model meteorology and not the underlying emissions inventory in winter months when
770 looking at individual observing sites. Not only do certain synoptic setups allow for minimum
771 absolute errors in the predicted values, but the timing and location of frontal passages can
772 significantly impact the model performance at predicting CO₂ mole fractions. We also find that
773 the errors associated with atmospheric transport are not restricted to certain times of day. This
774 suggests that filtering data based on model performance rather than time of day (such as using
775 only mid-afternoon observations) for atmospheric inversions might yield better overall results.
776 Thus, further methods, such as machine learning algorithms, are needed to better identify time
777 periods where the simulated transport performs well. To improve the simulated CO₂ mole
778 fractions error relative to observations, the prediction of key meteorological variables such as
779 wind speed and direction and the height of the PBL must be improved, either through more
780 advanced physics schemes or through data assimilation techniques. As such, minimizing errors
781 associated with atmospheric transport and dispersion generally will improve the performance of
782 estimated fossil fuel CO₂ emissions more than improving emission priors.

783 **Acknowledgments**

784 We acknowledge support for this project from the FLAGG-MD grant from NIST's Greenhouse
785 Gas Measurements program (Cooperative Agreement #70NANB14H333). We would like to
786 thank A. Andrews (NOAA) and S. de Wekker (UVA) for use of the in situ observations of CO₂
787 at SNP. We also would like to thank the team at Earth Networks for their help and support.
788 Certain commercial equipment, instruments, or materials are identified in this paper in order to
789 specify the experimental procedure adequately. Such identification is not intended to imply
790 recommendation or endorsement by the National Institute of Standards and Technology, nor is it
791 intended to imply that the materials or equipment identified are necessarily the best available for
792 the purpose. The observational data for SNP are available from NOAA, and the observations
793 from the other sites, as well as the modeled time series, are available from NIST at
794 <http://data.nist.gov> (doi TBD).

795 **References**

- 796 Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A.,
797 Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M.,
798 Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier,
799 C. O., De Wekker, S. F. J., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO₂, CO, and CH₄
800 measurements from tall towers in the NOAA Earth System Research Laboratory's Global
801 Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and
802 recommendations for future high-accuracy greenhouse gas monitoring efforts, *Atmos. Meas.*
803 *Techn.*, 7, 647-687, 10.5194/amt-7-647-2014, 2014.
- 804 Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., Huang,
805 J., Elvidge, C., and Baugh, K.: A multiyear, global gridded fossil fuel CO₂ emission data
806 product: Evaluation and analysis of results, *Journal of Geophysical Research: Atmospheres*, 119,
807 10,213-210,231, 10.1002/2013JD021296, 2014.
- 808 Beck, V., T. Koch, R. Kretschmer, J. Marshall, R. Ahmadov, C. Gerbig, D. Pillai, and M.
809 Heimann, 2011: The WRF Greenhouse Gas Model (WRF-GHG) Technical Report No. 25, Max
810 Planck Institute for Biogeochemistry, Jena, Germany, available online at [http://www.bgc-](http://www.bgc-jena.mpg.de/bgc-systems/index.shtml)
811 [jenampg.de/bgc-systems/index.shtml](http://www.bgc-jena.mpg.de/bgc-systems/index.shtml)
- 812 Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J.,
813 Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J.,
814 Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric
815 chemistry models: current status and future prospects for coupled chemistry meteorology
816 models, *Atmos. Chem. Phys.*, 15, 5325-5358, <https://doi.org/10.5194/acp-15-5325-2015>, 2015.
- 817 Breon, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I., Ramonet, M.,
818 Dieudonne, E., Lopez, M., Schmidt, M., Perrussel, O., and Ciais, P.: An attempt at estimating
819 Paris area CO₂ emissions from atmospheric concentration measurements, *Atmospheric*
820 *Chemistry and Physics*, 15, 1707-1724, 10.5194/acp-15-1707-2015, 2015.
- 821 Briber, B., Hutyra, L., Dunn, A., Raciti, S., and Munger, J.: Variations in Atmospheric CO₂
822 Mixing Ratios across a Boston, MA Urban to Rural Gradient, *Land*, 2, 304, 2013.
- 823 Carbontracker Team: Compilation of near real time atmospheric carbon dioxide data provided by
824 NOAA and EC; `obspace_co2_1_NRT_v3.3_2017-04-19`; NOAA Earth System Research
825 Laboratory, Global Monitoring Division. <http://doi.org/10.15138/G3G01J>, 2017.
- 826 Chou, M.-D., Suarez, M. J., Liang, X.-Z., Yan, M. M.-H., and Cote, C.: A thermal infrared
827 radiation parameterization for atmospheric studies, 2001.
- 828 Ciais P, Rayner P, Chevallier F, Bousquet P, Logan M, Peylin P, Ramonet M (2010)
829 Atmospheric inversions for estimating CO₂ fluxes: methods and perspectives. *Clim Change*
830 103(1-2):69-92. <https://doi.org/10.1007/s10584-010-9909-3>
- 831 Clarke L., K. Jiang, K. Akimoto, M. Babiker, G. Blanford, K. Fisher-Vanden, J.-C. Hourcade, V.
832 Krey, E. Kriegler, A. Löschel, D. McCollum, S. Paltsev, S. Rose, P.R. Shukla, M. Tavoni,
833 B.C.C. van der Zwaan, and D.P. van Vuuren: Assessing Transformation Pathways. In: *Climate*
834 *Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth*
835 *Assessment Report of the Intergovernmental Panel on Climate Change* [Edenhofer, O., R. Pichs-
836 Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P.

837 Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C.
838 Minx (eds.)). Cambridge University Press, Cambridge, United Kingdom and New York, NY,
839 USA, 2014.

840 Deng, A., T. Lauvaux, K. J. Davis, B. J. Gaudet, N. Miles, S. J. Richardson, K. Wu, D. P.
841 Sarmiento, R. M. Hardesty, and T. A. Bonin (2017), Toward reduced transport errors in a high
842 resolution urban CO₂ inversion system, *Elem Sci Anth*, 5.

843 Dlugokencky, E. J., Myers, R. C., Lang, P. M., Masarie, K. A., Crotwell, A. M., Thoning, K. W.,
844 Hall, B. D., Elkins, J. W., and Steele, L. P.: Conversion of NOAA atmospheric dry air CH₄ mole
845 fractions to a gravimetrically prepared standard scale, *Journal of Geophysical Research:*
846 *Atmospheres*, 110, doi:10.1029/2005JD006035, 2005.

847 Etheridge, D. M., L. P. Steele, R. L. Langenfelds, R. J. Francey, J.-M. Barnola, and V. I. Morgan
848 (1996), Natural and anthropogenic changes in atmospheric CO₂ over the last 1000 years from air
849 in Antarctic ice and firn, *J. Geophys. Res.*, 101(D2), 4115–4128, doi: 10.1029/95JD03410.

850 Feng, S., Lauvaux, T., Newman, S., Rao, P., Ahmadov, R., Deng, A. J., Diaz-Isaac, L. I., Duren,
851 R. M., Fischer, M. L., Gerbig, C., Gurney, K. R., Huang, J. H., Jeong, S., Li, Z. J., Miller, C. E.,
852 O’Keeffe, D., Patarasuk, R., Sander, S. P., Song, Y., Wong, K. W., and Yung, Y. L.: Los Angeles
853 megacity: a high-resolution land-atmosphere modelling system for urban CO₂ emissions,
854 *Atmospheric Chemistry and Physics*, 16, 9019-9045, 10.5194/acp-16-9019-2016, 2016.

855 Fischer, M. L., Parazoo, N., Brophy, K., Cui, X., Jeong, S., Liu, J., Keeling, R., Taylor, T. E.,
856 Gurney, K., Oda, T., and Graven, H.: Simulating estimation of California fossil fuel and
857 biosphere carbon dioxide exchanges combining in situ tower and satellite column observations,
858 *Journal of Geophysical Research: Atmospheres*, 122, 3653-3671, 10.1002/2016JD025617, 2017.

859 Gately, C. K., & Hutyra, L. R. (2017). Large uncertainties in urban-scale carbon emissions.
860 *Journal of Geophysical Research: Atmospheres*, 122, 11,242–11,260.
861 <https://doi.org/10.1002/2017JD027359>

862 Gemmill, W., Katz, B., and Li, X.: Daily real-time global sea surface temperature-high
863 resolution analysis at NOAA/NCEP, NCEP Off. Note, 260, 39, 2007.

864 Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and
865 Eder, B.: Fully coupled "online" chemistry within the WRF model, *Atmospheric Environment*,
866 39, 6957-6975, 10.1016/j.atmosenv.2005.04.027, 2005.

867 Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., and de
868 la Rue du Can, S.: High Resolution Fossil Fuel Combustion CO₂ Emission Fluxes for the United
869 States, *Environmental Science & Technology*, 43, 5535-5541, 10.1021/es900806c, 2009.

870 Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A Revised Approach to Ice Microphysical Processes
871 for the Bulk Parameterization of Clouds and Precipitation, *Monthly Weather Review*, 132, 103-
872 120, 10.1175/1520-0493(2004)132<0103:aratim>2.0.co;2, 2004.

873 Hong, S.-Y., Noh, Y., and Dudhia, J.: A New Vertical Diffusion Package with an Explicit
874 Treatment of Entrainment Processes, *Monthly Weather Review*, 134, 2318-2341,
875 10.1175/mwr3199.1, 2006.

876 Hutchins, M. G., Colby, J. D., Marland, G., and Marland, E.: A comparison of five high-
877 resolution spatially-explicit, fossil-fuel, carbon dioxide emission inventories for the United

878 States, Mitigation and Adaptation Strategies for Global Change, 1-26, 10.1007/s11027-016-
879 9709-9, 2016.

880 IEA, 2012. CO₂ emissions from fuel combustion 1971–2010, 2012 Edition, Paris: International
881 Energy Agency (IEA), 2012.

882

883 Intergovernmental Panel on Climate Change (2013). Climate Change 2013: The Physical
884 Science Basis. Contribution of Working Group I to the
885 Fifth Assessment Report of the Intergovernmental Panel on Climate Change, In T. F. Stocker, et
886 al. (Eds.), (1535 pp.). Cambridge, U. K., and New York: Cambridge University Press.
887 <https://doi.org/10.1017/CBO9781107415324>

888 Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, *Journal of Applied*
889 *Meteorology*, 43, 170-181, 10.1175/1520-0450(2004)043<0170:tkcpau>2.0.co;2, 2004.

890 Kort, E. A., Angevine, W. M., Duren, R., and Miller, C. E.: Surface observations for monitoring
891 urban fossil fuel CO₂ emissions: Minimum site location requirements for the Los Angeles
892 megacity, *Journal of Geophysical Research-Atmospheres*, 118, 1-8, 10.1002/jgrd.50135, 2013.

893 Kretschmer, R., C. Gerbig, U. Karstens, G. Biavati, A. Vermeulen, F. Vogel, S. Hammer, and K.
894 U. Totsche: Impact of Optimized Mixing Heights on Simulated Regional Atmospheric Transport
895 of CO₂, *Atmos. Chem. Phys.* 14, no. 14 (July 16, 2014): 7149-72. [https://doi.org/10.5194/acp-14-](https://doi.org/10.5194/acp-14-7149-2014)
896 [7149-2014](https://doi.org/10.5194/acp-14-7149-2014), 2014.

897 Lauvaux, T., N.L. Miles, S.J. Richardson, A. Deng, D.R. Stauffer, K.J. Davis, G. Jacobson, C.
898 Rella, G. Calonder, and P.L. DeCola, 2013: Urban Emissions of CO₂ from Davos, Switzerland:
899 The First Real-Time Monitoring System Using an Atmospheric Inversion Technique. *J. Appl.*
900 *Meteor. Climatol.*, 52, 2654–2668, <https://doi.org/10.1175/JAMC-D-13-038.1>

901 Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet,
902 B., Gurney, K. R., Huang, J., and O'Keefe, D.: High-resolution atmospheric inversion of urban
903 CO₂ emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX),
904 *Journal of Geophysical Research: Atmospheres*, 2016.

905 Lee, T. R., De Wekker, S. F. J., Andrews, A. E., Kofler, J., and Williams, J.: Carbon dioxide
906 variability during cold front passages and fair weather days at a forested mountaintop site,
907 *Atmospheric Environment*, 46, 405-416, <https://doi.org/10.1016/j.atmosenv.2011.09.068>, 2012.

908 Leip, A., Skiba, U., Vermeulen, A., and Thompson, R. L.: A complete rethink is needed on how
909 greenhouse gas emissions are quantified for national reporting, *Atmospheric Environment*, 174,
910 237-240, <https://doi.org/10.1016/j.atmosenv.2017.12.006>, 2018.

911 Locatelli, R., P. Bousquet, M. Saunois, F. Chevallier, and C. Cressot: Sensitivity of the Recent
912 Methane Budget to LMDz Sub-Grid-Scale Physical Parameterizations, *Atmospheric Chemistry*
913 *and Physics* 15, no. 17 (September 1, 2015): 9765-80. <https://doi.org/10.5194/acp-15-9765-2015>,
914 2015.

915 Lopez-Coto, I., Ghosh, S., Prasad, K., and Whetstone, J.: Tower-based greenhouse gas
916 measurement network design—The National Institute of Standards and Technology North East
917 Corridor Testbed, *Advances in Atmospheric Sciences*, 34, 1095-1105, 10.1007/s00376-017-
918 6094-6, 2017.

919 McDonald, B. C., Z. C. McBride, E. W. Martin, and R. A. Harley (2014), High-resolution
920 mapping of motor vehicle carbon dioxide emissions, *J. Geophys. Res. Atmos.*, 119, 5283–5298,
921 doi: 10.1002/2013JD021219.

922 McKain, K., Wofsy, S. C., Nehr Korn, T., Eluszkiewicz, J., Ehleringer, J. R., and Stephens, B. B.:
923 Assessment of ground-based atmospheric observations for verification of greenhouse gas
924 emissions from an urban region, *Proceedings of the National Academy of Sciences of the United*
925 *States of America*, 109, 8423-8428, 10.1073/pnas.1116645109, 2012.

926 McKain, K., Down, A., Raciti, S. M., Budney, J., Hutyra, L. R., Floerchinger, C., Herndon, S.
927 C., Nehr Korn, T., Zahniser, M. S., Jackson, R. B., Phillips, N., and Wofsy, S. C.: Methane
928 emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts,
929 *Proceedings of the National Academy of Sciences*, 112, 1941-1946, 10.1073/pnas.1416261112,
930 2015.

931 McRae, J. E., and Graedel, T. E.: Carbon Dioxide in the Urban Atmosphere: Dependencies and
932 Trends, *J. Geophys. Res.-Oceans*, 84, 5011-5017, 10.1029/JC084iC08p05011, 1979.

933 Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D.,
934 Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H.,
935 Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American Regional Reanalysis, *Bulletin of*
936 *the American Meteorological Society*, 87, 343-360, 10.1175/bams-87-3-343, 2006.

937 Miles, N. L., Richardson, S. J., Lauvaux, T., Davis, K. J., Balashov, N. V., Deng, A., Turnbull, J.
938 C., Sweeney, C., Gurney, K. R., and Patarasuk, R.: Quantification of urban atmospheric
939 boundary layer greenhouse gas dry mole fraction enhancements in the dormant season: Results
940 from the Indianapolis Flux Experiment (INFLUX), *Elem Sci Anth*, 5, 2017.

941 Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer
942 for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave,
943 *Journal of Geophysical Research: Atmospheres*, 102, 16663-16682, 10.1029/97JD00237, 1997.

944 Mueller, K., Yadav, V., Lopez-Coto, I., Karion, A., Gourdji, S., Martin, C., & Whetstone,
945 J. (2018). Siting background towers to characterize incoming air for urban greenhouse gas
946 estimation: A case study in the Washington, DC/Baltimore area. *Journal of Geophysical*
947 *Research: Atmospheres*, 123. <https://doi.org/10.1002/2017JD027364>

948 Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R., and Deng, F.:
949 Improving the temporal and spatial distribution of CO₂ emissions from global fossil fuel
950 emission data sets, *Journal of Geophysical Research: Atmospheres*, 118, 917-933,
951 10.1029/2012JD018196, 2013.

952 Nehr Korn, T., Henderson, J., Leidner, M., Mountain, M., Eluszkiewicz, J., McKain, K., and
953 Wofsy, S.: WRF simulations of the urban circulation in the Salt Lake City area for CO₂
954 modeling, *Journal of Applied Meteorology and Climatology*, 52, 323-340, 2013.

955 National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory,
956 Global Monitoring Division, Trends in Atmospheric Carbon Dioxide,
957 <https://www.esrl.noaa.gov/gmd/ccgg/trends/full.html>, Last accessed: January 5, 2018.

958 Oda, T., and Maksyutov, S.: A very high-resolution (1 km×1 km) global fossil fuel CO₂
959 emission inventory derived using a point source database and satellite observations of nighttime
960 lights, *Atmos. Chem. Phys.*, 11, 543-556, 10.5194/acp-11-543-2011, 2011.

961 Oda, T., and Maksyutov, S. (2015), ODIAC Fossil Fuel CO₂ Emissions Dataset (Version name:
962 ODIAC2016), Center for Global Environmental Research, National Institute for Environmental
963 Studies, doi:10.17595/20170411.001. (Reference date : 2017/02/02)

964 Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic
965 Carbon dioxide (CO₂), version 2016 (ODIAC2016): A global monthly fossil fuel CO₂ gridded
966 emissions data product for tracer transport simulations and surface flux inversions, *Earth Syst.*
967 *Sci. Data*, 10, 87-107, <https://doi.org/10.5194/essd-10-87-2018>, 2018

968 Olivier, J. G. J., Van Aardenne, J. A., Dentener, F. J., Pagliari, V., Ganzeveld, L. N., and Peters,
969 J. A. H. W.: Recent trends in global greenhouse gas emissions: regional trends 1970–2000 and
970 spatial distribution of key sources in 2000, *Environmental Sciences*, 2, 81-99,
971 10.1080/15693430500400345, 2005.

972 Parazoo, N. C., Denning, A. S., Kawa, S. R., Corbin, K. D., Lokupitiya, R. S., and Baker, I. T.:
973 Mechanisms for synoptic variations of atmospheric CO₂ in North America, South America and
974 Europe, *Atmos. Chem. Phys.*, 8, 7239-7254, <https://doi.org/10.5194/acp-8-7239-2008>, 2008.

975 Poulida, O., R. R. Dickerson, B. G. Doddridge, J. Z. Holland, R. G. Wardell, and J. G. Watkins:
976 Trace Gas Concentrations and Meteorology in Rural Virginia: 1. Ozone and Carbon-Monoxide,
977 *Journal of Geophysical Research-Atmospheres*, 96(D12), 22461-22475, 1991.

978 Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J.
979 B., Bruhwiler, L. M., Petron, G., and Hirsch, A. I.: An atmospheric perspective on North
980 American carbon dioxide exchange: CarbonTracker, *Proceedings of the National Academy of*
981 *Sciences*, 104, 18925-18930, 2007.

982 Rayner, P. J., Raupach, M. R., Paget, M., Peylin, P., and Koffi, E.: A new global gridded data set
983 of CO₂ emissions from fossil fuel combustion: Methodology and evaluation, *Journal of*
984 *Geophysical Research: Atmospheres*, 115, n/a-n/a, 10.1029/2009JD013439, 2010.

985 Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T.,
986 Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Dool, H. v. d., Zhang, Q.,
987 Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, *Journal of*
988 *Climate*, 27, 2185-2208, 10.1175/jcli-d-12-00823.1, 2014.

989 Sargent, M., Barrera, Y., Nehr Korn, T., Hutyra, L. R., Gately, C. K., Jones, T., McKain, K.,
990 Sweeney, C., Hegarty, J., Hardiman, B., and Wofsy, S. C.: Anthropogenic and biogenic CO₂
991 fluxes in the Boston urban region, *Proceedings of the National Academy of Sciences*,
992 10.1073/pnas.1803715115, 2018.

993 Shusterman, A. A., Teige, V., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The
994 BErkeley Atmospheric CO₂ Observation Network: initial evaluation, *Atmos. Chem. Phys.*
995 *Discuss.*, 2016, 1-23, 10.5194/acp-2016-530, 2016.

996 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-
997 Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3,
998 2008.

999 Strong, C., C. Stwertka, D. R. Bowling, B. B. Stephens, and J. R. Ehleringer, Urban carbon
1000 dioxide cycles within the Salt Lake Valley: A multiple-box model validated by observations, *J.*
1001 *Geophys. Res.*, 116, D15307, doi: 10.1029/2011JD015693, 2011.

1002 Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G.,
1003 Wegiel, J., and Cuenca, R.: Implementation and verification of the unified NOAA land surface
1004 model in the WRF model, 20th conference on weather analysis and forecasting/16th conference
1005 on numerical weather prediction, 2004.

1006 Thiébaux, J., Rogers, E., Wang, W., and Katz, B.: A new high-resolution blended real-time
1007 global sea surface temperature analysis, *Bulletin of the American Meteorological Society*, 84,
1008 645-656, 2003.

1009 Turnbull, J. C., Sweeney, C., Karion, A., Newberger, T., Lehman, S. J., Tans, P. P., Davis, K. J.,
1010 Lauvaux, T., Miles, N. L., Richardson, S. J., Cambaliza, M. O., Shepson, P. B., Gurney, K.,
1011 Patarasuk, R., and Razlivanov, I.: Toward quantification and source sector identification of fossil
1012 fuel CO₂ emissions from an urban area: Results from the INFLUX experiment, *Journal of*
1013 *Geophysical Research-Atmospheres*, 120, 292-312, 10.1002/2014jd022555, 2015.

1014 United Nations: Cities and Climate Change: Global Report on Human Settlements 2011. 2011.
1015 Available online: [https://unhabitat.org/?mbt_book=cities-and-climate-change-global-report-on-](https://unhabitat.org/?mbt_book=cities-and-climate-change-global-report-on-human-settlements-2011)
1016 [human-settlements-2011](https://unhabitat.org/?mbt_book=cities-and-climate-change-global-report-on-human-settlements-2011) (accessed on 15 September 2018).

1017 Verhulst, K. R., Karion, A., Kim, J., Salameh, P. K., Keeling, R. F., Newman, S., Miller, J.,
1018 Sloop, C., Pongetti, T., Rao, P., Wong, C., Hopkins, F. M., Yadav, V., Weiss, R. F., Duren, R.
1019 M., and Miller, C. E.: Carbon dioxide and methane measurements from the Los Angeles
1020 Megacity Carbon Project – Part 1: calibration, urban enhancements, and uncertainty estimates,
1021 *Atmos. Chem. Phys.*, 17, 8313-8341, 10.5194/acp-17-8313-2017, 2017.

1022 Vogel, F. R., Frey, M., Stauffer, J., Hase, F., Broquet, G., Xueref-Remy, I., Chevallier, F., Ciais,
1023 P., Sha, M. K., Chelin, P., Jeseck, P., Janssen, C., Te, Y.-V., Groß, J., Blumenstock, T., Tu, Q.,
1024 and Orphal, J.: XCO₂ in an emission hot-spot region: the COCCON Paris campaign 2015,
1025 *Atmos. Chem. Phys. Discuss.*, <https://doi.org/10.5194/acp-2018-595>, in review, 2018.

1026 Whelpdale, D. M., Low, T. B., and Kolomeychuk, R. J.: Advection climatology for the east coast
1027 of North America, *Atmospheric Environment (1967)*, 18, 1311-1327, 10.1016/0004-
1028 6981(84)90040-4, 1984.

1029 Wunch, D., G. C. Toon, P. O. Wennberg, S. C. Wofsy, B. Stephens, M. L. Fisher, O. Uchino, J.
1030 B. Abshire, P. F. Bernath, S. C. Biraud, J.-F. L. Blavier, C. D. Boone, K. P. Bowman, E. V.
1031 Browell, T. Campos, B. J. Connor, B. C. Daube, N. M. Deutscher, M. Diao, J. W. Elkins, C.
1032 Gerbig, E. Gottlieb, D. W. T. Griffith, D. F. Hurst, R. Jiménez, G. Keppel-Aleks, E. A. Kort, R.
1033 Macatangay, T. Machida, H. Matsueda, F. L. Moore, I. Morino, S. Park, J. Robinson, C. M.
1034 Roehl, Y. Sawa, V. Sherlock, C. Sweeney, T. Tanaka, and M. A. Zondlo (2010), Calibration of
1035 the Total Carbon Column Observing Network using aircraft profile data, *Atmospheric*
1036 *Measurement Techniques*, 3(5), 1351-1362, doi:10.5194/amt-3-1351-2010.

1037 Zeng, N., Mariotti, A., and Wetzal, P.: Terrestrial mechanisms of interannual CO₂ variability,
1038 *Global Biogeochemical Cycles*, 19, 2005.

1039 Zhao, C. L., and Tans, P. P.: Estimating uncertainty of the WMO mole fraction scale for carbon
1040 dioxide in air, *Journal of Geophysical Research: Atmospheres*, 111, doi:10.1029/2005JD006003,
1041 2006.

1 Supporting Information for

2 **Evaluation of WRF-Chem Simulated Carbon Dioxide Atmospheric Transport and Emissions in the**
3 **Baltimore-Washington Metropolitan Area**

4 Cory R. Martin¹, N. Zeng^{1,2}, A. Karion³, K. Mueller³, S. Ghosh⁴, I. Lopez-Coto⁴, K. R. Gurney⁵, T. Oda^{6,7},
5 K. Prasad⁴, Y. Liu^{1*}, R. R. Dickerson^{1,2}, and J. Whetstone³

6
7 ¹Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA

8 ²Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

9 ³Special Programs Office, National Institute of Standards and Technology, Gaithersburg, MD, USA

10 ⁴Fire Research Division, National Institute of Standards and Technology, Gaithersburg, MD, USA

11 ⁵School of Life Sciences, Arizona State University, Tempe, AZ, USA

12 ⁶Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

13 ⁷Goddard Earth Sciences Research and Technology, Universities Space Research Association, Columbia, MD, USA

14 *Now at: College of Geosciences, Texas A&M University, College Station, TX, USA

15
16
17 **Contents of this file**

18
19 Text S1 to S2

20 Figure S1 to S2

21 Table S1 to S2

22
23 **Introduction**

24 Text S1 describes the evaluation of the model meteorology using observations of 10 m wind speed
25 and direction as well as estimated heights of the planetary boundary layer.

26 Text S2 describes the linear regression of the model tracers described in Sect. 3.1 and shown in Fig.
27 4.

28 Figure S1 shows the mean and spread of the difference between WRF and observations for 10 m wind
29 speed, 10 m wind direction, and height of the planetary boundary layer.

30 Figure S2. Two-day running correlations of the observed timeseries vs each of the modeled CO₂
31 timeseries.

32 Table S1 provides a summary of the three WRF domains and their configurations

33 Table S2 shows the statistics of the linear regression described in Sect. 3.1 and Text S2.

34
35
36

1 **Text S1.**

2 Ten Automated Surface Observing System (ASOS) or Automated Weather Observing System
3 (AWOS) surface observation stations are used to evaluate WRF's prediction of near-surface wind speed
4 and direction. The datasets were downloaded from the National Weather Service's Meteorological
5 Assimilation Data Ingest System (MADIS) and processed by the supplied API. The observations for wind
6 are generally sited at 10m above ground level, and the WRF interpolated 10m winds are used for the
7 comparison. Both the observations and the model output provide the U and V components of the wind
8 vector, which are converted to speed and degrees from north for subsequent comparison.
9

10 To evaluate WRF's simulation of the vertical temperature profile of the atmosphere as well as its
11 calculation of the PBL, observations from the Aircraft Communications Addressing and Reporting
12 System (ACARS) are used. Available from the MADIS API, profiles are created when certain aircraft
13 take off and land at airports. Two of the three airports (KBWI, Baltimore-Washington Thurgood Marshall
14 International and KDCA, Reagan National) located within the NEC-B/W regions have sufficient data
15 available for February 2016. The WRF PBL scheme calculates the height of the PBL as an output
16 variable, but the ACARS dataset only provides vertical profiles of observed quantities such as
17 temperature and pressure. To compare, an algorithm is used to estimate the observed and modeled PBL
18 height by computing the vertical potential temperature gradient and finding where the gradient is at its
19 relative maximum. For each hour, all available ACARS profiles meeting the following criteria are used:
20 the lowest reported data point must be below 300 m above ground level, there must be at least 10 data
21 points, and the computed PBL height must be below 2100 m (in case the tropopause is detected by the
22 algorithm). The PBL height is computed from all valid profiles, and then the mean of these computed
23 heights is used. Any hour containing fewer than three individual profiles is not used. To supplement the
24 ACARS profile data at the third airport (KIAD; Washington-Dulles International), the radiosonde profiles
25 launched by the National Weather Service in Sterling, VA (KLWX) are used. The office, located onsite at
26 the airport, launches radiosondes typically twice per day (0000 UTC and 1200 UTC; 7pm and 7am local
27 time), so the temporal coverage of observed profiles is sparse compared to the other two airports.
28
29

30 For all hours, WRF has a positive wind speed bias at all but one site (KDCA, Reagan National
31 Airport) with an average over all ten locations of 1.2 m/s. This slight positive bias in wind speed is
32 consistent with previous comparisons of WRF to observations in similar work (e.g. Nehr Korn et al., 2012;
33 Feng et al., 2016). The average wind direction bias over all sites and for the entire month is approximately
34 2.8°. The standard deviation of the average difference between model and observations is virtually the
35 same as the mean bias for wind speed, with an average standard deviation of 1.2 m/s to one decimal place.
36

37 When averaged over the entire evaluation period, the comparisons between estimated PBL heights
38 from WRF and the observations at each airport exhibit different behavior. Both the computed PBL height
39 from the potential temperature gradient and the YSU predicted PBL height (using the Richardson
40 number) are compared to the calculated height from the observed profiles. KBWI is the best performer,
41 with a mean error of the computed height of 11 m, compared to an error of -145 m for the YSU predicted
42 height. At KDCA, the YSU predicted height is 30 m too high on average, whereas using the potential
43 temperature profile results in a higher bias of 325 m. Finally, at KLWX, using only the radiosondes, the
44 mean difference over the month is negative for both methods, with an average of -118 m below
45 observations for the computed height and -481 m for the YSU provided height. For most days, KLWX
46 only has profiles twice a day, and as such may not be representative of the model's overall performance.
47 Overall, the YSU estimated PBL height is an average of -71 m from the observations, and the potential
48 temperature profile method results in a 150 m high bias. The aforementioned meteorological fields from
49 WRF agree well with observations on average, but can vary greatly for any given hour. See Fig. S11 for
50 time series of the mean (solid line) and range (min and max; shaded regions) difference between WRF

1 and observations for all three variables: PBL height (blue is potential temperature; red YSU), wind speed,
2 and wind direction.

3

4 **Text S2.**

5 For the regression described in Section 3.1 and shown in Figure 4, an ordinary least squares linear
6 regression was employed. The linear model can be written as the following:

7

8

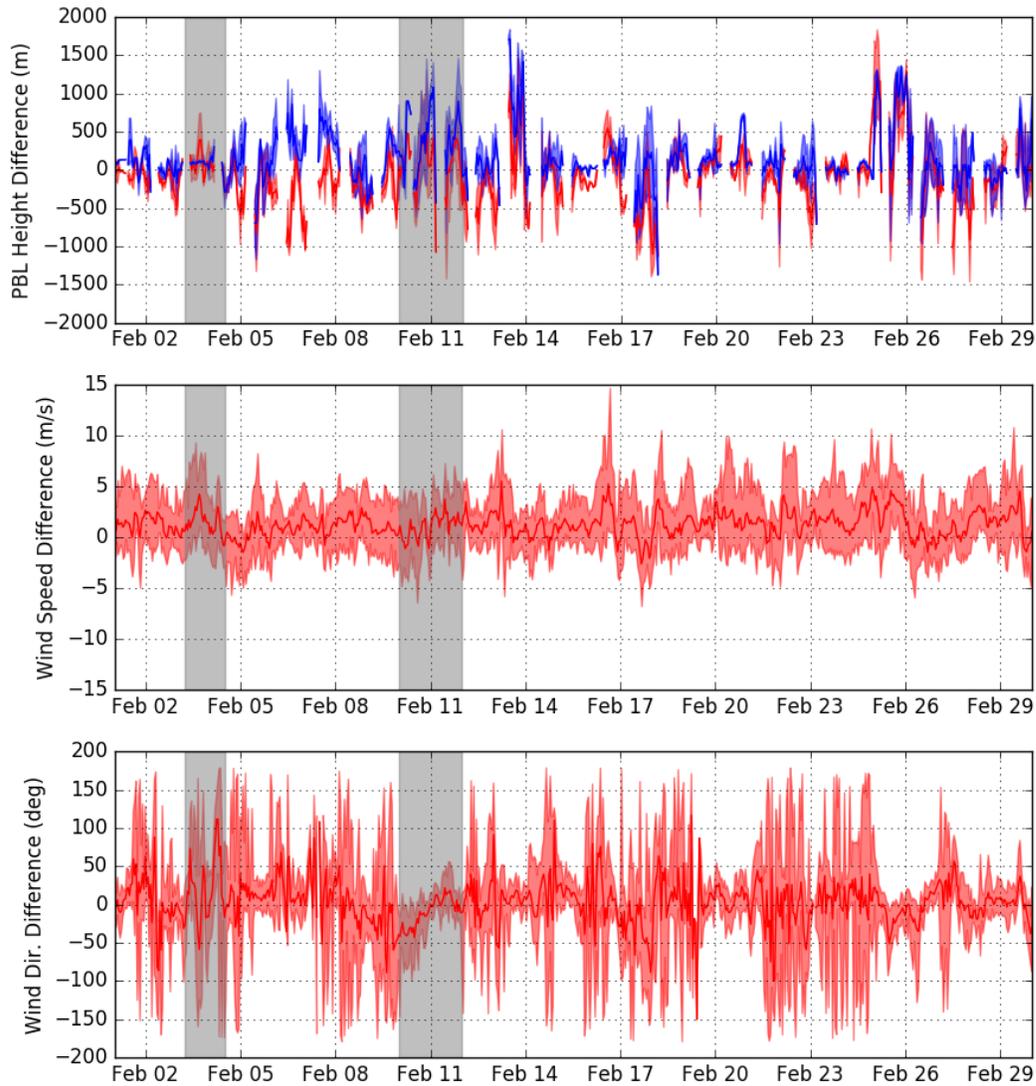
$$y = \alpha + \beta x + \epsilon$$

9

10 where: the observed CO₂ mole fractions is the independent variable, x , the CO₂ mole fractions simulated
11 by WRF-Chem is the dependent variable, y , α is the intercept of the line, β is the slope, and ϵ is the
12 residual, the difference between the actual dependent variable and the prediction of the linear model. For
13 each of the simulated CO₂ time series, first the data are filtered by removing any outliers (in this case
14 points where the difference between the observed and simulated CO₂ is more than three times the
15 standard deviation) and then a linear regression is fitted.

15

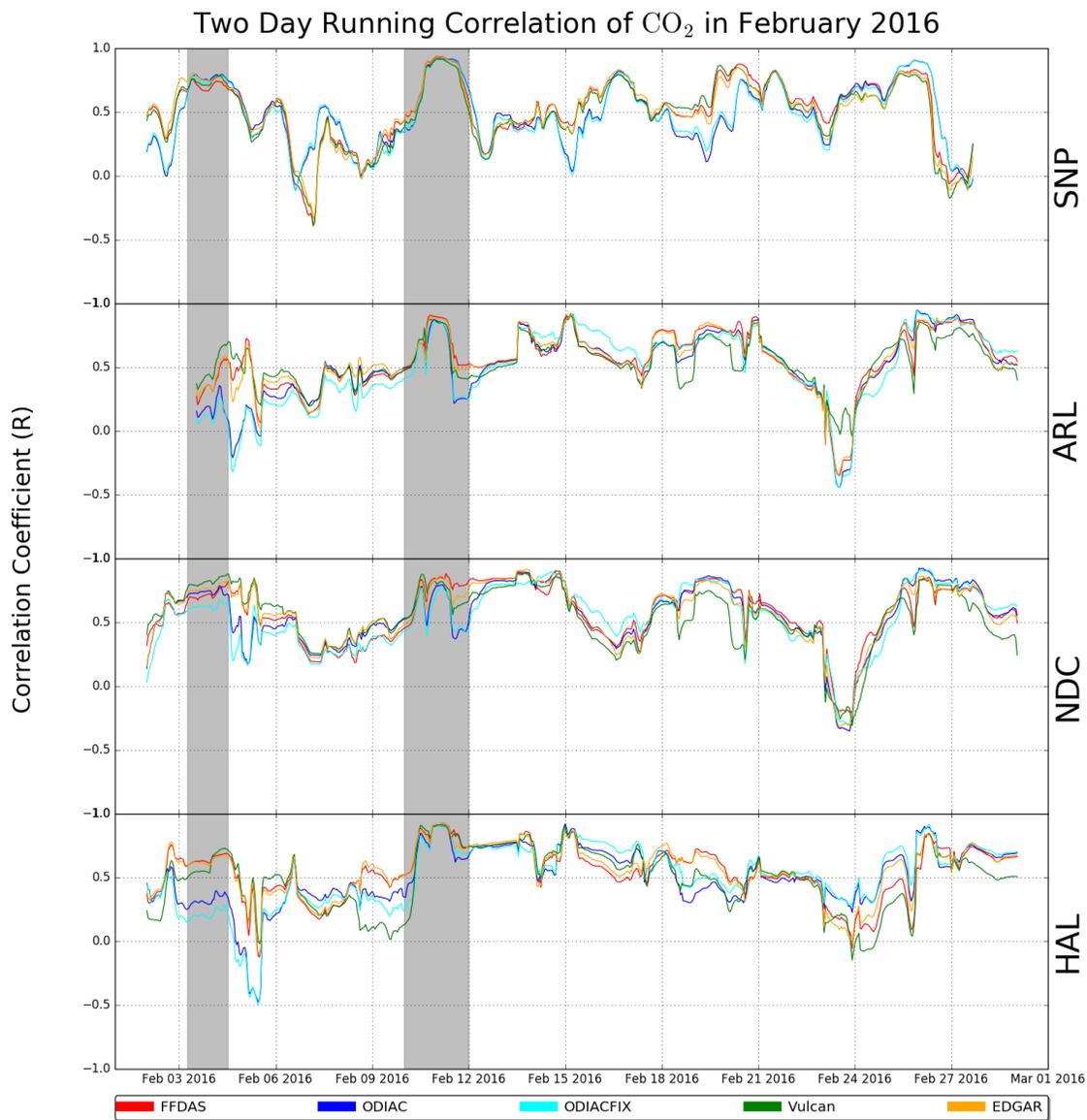
February 2016 WRF-Obs. Met Evaluation



1

2 **Figure S1.** Mean (solid line) and spread (shaded area between minimum and maximum values) of the
3 difference between WRF and observations for: PBL height (YSU computed PBL height in red; potential
4 temperature profile computed in blue), wind speed, and wind direction. For PBL, observations are for
5 three airports (KDCA, KBWI, KIAD), and for wind observations, ten sites in domain d03 are used.

1 Positive values indicate a larger quantity from the model. The two cases are shaded in gray, as in Figs. 3
2 and 4.



3
4 **Figure S2.** Two-day running correlations of the observed timeseries vs each of the modeled CO₂
5 timeseries (each color for the corresponding anthropogenic emissions inventory used) for each of the four
6 observing sites (the four panels).

7
8

Domain	d01	d02	d03
Horizontal Resolution	9 km	3 km	1 km
Vertical Levels	50 (from surface to 50 hPa)		
Microphysics	Single-Moment 5-class		
Radiation	RRTM longwave and Goddard shortwave		
Cumulus Parameterization	Kain-Fritsch	Kain-Fritsch	None
PBL Scheme	Yonsei University		
Land Surface	Noah Land Surface Model		

1
2
3
4

Table S1. Summary of the three WRF domains and their configurations

Site Code	Tracer Source	Inlet				RMSE		# outliers	MAE
			Bias	Slope	Intercept	R ²			
ARL	FFDAS	50m	1.47	0.77	98.95	0.79	7.26	200	4.87
ARL	ODIAC	50m	-3.54	0.58	174.28	0.14	6.54	167	5.06
ARL	ODIACFIX	50m	-2.27	0.59	170.80	0.38	6.96	171	5.23
ARL	Vulcan	50m	-1.34	0.61	162.78	0.53	6.47	174	4.66
ARL	EDGAR	50m	-3.25	0.58	172.84	0.21	6.38	185	4.89
ARL	FFDAS	92m	-0.93	0.74	109.54	0.76	5.61	160	3.97
ARL	ODIAC	92m	-4.54	0.49	208.79	0.31	6.42	143	5.07
ARL	ODIACFIX	92m	-3.78	0.55	186.30	0.03	6.36	158	4.97
ARL	Vulcan	92m	-2.80	0.58	175.44	0.29	5.61	174	4.10
ARL	EDGAR	92m	-4.35	0.50	204.97	0.24	6.15	149	4.77
HAL	FFDAS	29m	5.25	0.88	56.62	0.42	11.43	181	7.56
HAL	ODIAC	29m	-2.40	0.65	146.57	0.48	8.24	190	6.19
HAL	ODIACFIX	29m	-0.52	0.81	78.91	0.87	9.06	195	6.67
HAL	Vulcan	29m	1.08	0.79	87.94	0.86	9.67	204	6.73
HAL	EDGAR	29m	-1.47	0.66	142.61	0.61	8.18	187	6.00
HAL	FFDAS	58m	4.52	0.90	44.30	0.18	8.98	174	6.15
HAL	ODIAC	58m	-1.34	0.67	137.63	0.63	7.12	171	5.34
HAL	ODIACFIX	58m	0.26	0.78	94.16	0.84	7.72	190	5.69
HAL	Vulcan	58m	1.23	0.77	97.68	0.80	7.99	201	5.54
HAL	EDGAR	58m	-1.04	0.64	152.35	0.61	6.58	178	4.86
NDC	FFDAS	45m	3.84	0.81	84.91	0.53	8.58	198	5.87
NDC	ODIAC	45m	-2.84	0.56	182.13	0.28	6.96	157	5.31
NDC	ODIACFIX	45m	-1.25	0.64	151.44	0.58	7.41	177	5.59
NDC	Vulcan	45m	0.00	0.69	129.66	0.73	7.16	190	5.18
NDC	EDGAR	45m	-2.83	0.58	173.72	0.30	6.84	188	5.24
NDC	FFDAS	91m	1.56	0.67	141.20	0.64	7.11	180	4.78
NDC	ODIAC	91m	-3.06	0.54	191.69	0.07	6.13	164	4.74
NDC	ODIACFIX	91m	-2.11	0.56	181.34	0.30	6.39	164	4.85
NDC	Vulcan	91m	-0.93	0.50	207.83	0.35	6.25	158	4.40
NDC	EDGAR	91m	-2.97	0.48	217.60	0.00	6.11	169	4.69
SNP	FFDAS	17m	-0.26	0.51	200.36	0.43	2.19	116	1.72
SNP	ODIAC	17m	-0.68	0.52	196.53	0.32	2.37	131	1.86
SNP	ODIACFIX	17m	-0.50	0.54	189.95	0.41	2.40	132	1.86
SNP	Vulcan	17m	-0.25	0.48	215.68	0.28	2.20	112	1.74
SNP	EDGAR	17m	-0.56	0.47	219.25	0.30	2.24	116	1.76
All	FFDAS	N/A	1.83	0.86	61.47	0.84	7.61	1716	4.94

All	ODIAC	N/A	-2.76	0.62	158.61	0.35	6.49	1425	4.79
All	ODIACFIX	N/A	-1.58	0.70	123.49	0.64	6.84	1498	4.95
All	Vulcan	N/A	-0.70	0.71	122.79	0.72	6.80	1524	4.65
All	EDGAR	N/A	-2.61	0.62	157.65	0.39	6.32	1516	4.63

1

2

3

4

5

6

7

8

9

10

Table S2. Results of a linear regression performed at each observing site and for both inlets where applicable for all five tracers as well as for each emissions inventory across all sites an inlet. For each regression, the observed dataset is used as the independent variable, x, and the model predicted values are used as the dependent variable, y. Provided in the table are the: bias before the regression (ppm), slope, intercept (ppm), the coefficient of determination (R^2), the root-mean-square error before the regression (RMSE; ppm), the number of outliers using the double median absolute deviation (out of 696 hours), and the mean absolute error before the regression (MAE; ppm), for each simulated time series dataset at each observing level.